

Informatie zoeken en vinden

Eric Sieverts

Instituut voor Media en Informatie Management

[versie 1.0 ; augustus 2006]

Of het nu in een werksituatie is of privé, mensen hebben voortdurend voor allerlei doeleinden informatie nodig. Soms heb je bepaalde heel praktische gegevens nodig. Hoe laat vertrekt morgenochtend een trein naar Maastricht? Waar kan ik deze camera het goedkoopst kopen? Vaak heb je ook veel inhoudelijker informatie nodig. Dat kan zijn om al bestaande kennis aan te vullen of te controleren, maar ook heel vaak om geïnformeerd te worden over nog tamelijk onbekende onderwerpen.

Als je op zoek gaat naar informatie moet je voortdurend keuzes maken. De keuze hoe je vraag af te bakenen, de keuze naar welk soort informatie je eigenlijk zoekt, de keuze waar je naar dat soort informatie gaat zoeken, de keuze hoe je vervolgens precies gaat zoeken. Voor sommige informatievragen – bijvoorbeeld voor die trein naar Maastricht – liggen de te maken keuzes nogal voor de hand. Voor andere zul je veel meer moeten weten over informatiesoorten, informatiebronnen, informatiecollecties, zoekhulpmiddelen en toe te passen zoekstrategieën. Deze tekst geeft een inleiding op deze onderwerpen, probeert hun onderlinge relatie aan te geven en laat zien hoe die verschillende aspecten in een zoekplan samenkomen.

Hoewel sinds de opkomst van het web enorm veel informatie digitaal via internet beschikbaar is, moet je je realiseren dat bibliotheken en mediatheken ook nog altijd heel veel informatie in papieren vorm bieden. Niettemin zal in deze tekst de nadruk liggen op digitale informatie, of liever gezegd op digitale zoekhulpmiddelen - ook voor die informatie "op papier". Anderzijds is lang niet alle belangrijke informatie op internet gratis beschikbaar. Als je verantwoord informatie wilt zoeken, is er dus heel wat meer dan alleen wat je met Google kunt vinden. Deze tekst heeft dus zeker ook betrekking op bronnen en zoeksystemen waar niet iedereen zomaar gratis bij kan.

Inhoud

1	Informatiesoorten en informatiebronnen	3
1.1.	Aard van de informatie	3
1.1.1.	Theorie.	3
1.1.2.	Onderzoeksresultaten.	3
1.1.3.	Bestuurlijke informatie.	4
1.1.4.	Gegevens, feiten, naslag.	4
1.1.5.	Nieuws.	4
1.2.	Aard van de informatiecollecties	5
1.2.1.	Primaire bronnen.	5
1.2.2.	Secundaire bronnen.	6
1.2.3.	Tertiaire bronnen.	7
1.3.	Aard van de inhoud van de informatiebron	8
1.3.1.	Bibliografische databases.	8
1.3.2.	Full-text databases	9
1.3.3.	Bestanden met feitelijke gegevens.	10
1.4.	Wel of niet op het web?	11
2	Informatiecollecties en hun zoekhulpmiddelen	13
3	Zoekstrategie en zoekmethode	16
3.1	De zoekvraag	16
3.2	Zoektechnieken	17
3.2.1	Best-match zoeken	17
3.2.2	Booleaans zoeken.	19
3.2.3	Truncatie en woordstammen	22
3.2.4	Zoekingangen; zoeken in velden	23
3.3	Zoekmethoden.	27
3.3.1	Vraaganalyse en zoekwoorden	28
3.3.2	Bouwsteenmethode	28
3.3.3	Best match zoeken	30
3.3.4	Uitbreiden van zoekresultaten.	31
3.3.5	Inperken van zoekresultaten	34
4	Zoekplan	36
4.1	Stap 1: Afbakening en aard van de vraag	36
4.2	Stap 2: Te gebruiken bronnen en systemen	36
4.3	Stap 3: Te gebruiken zoekwoorden.	37
4.4	Stap 4: Te gebruiken zoekmethode(s).	37
4.5	Stap 5: Beoordeling van de resultaten en selectie van de informatie.	38
4.6	Stap 6: Evaluatie van het zoektraject	38
4.7	Vastlegging van het zoektraject	38

1 Informatiesoorten en informatiebronnen

Om goed informatie te kunnen zoeken, moet je een duidelijk beeld hebben van het steeds complexere informatielandschap, waarin steeds meer informatie steeds vaker, in meer verschillende vormen, langs meer verschillende wegen beschikbaar is. Dat zal een algemeen beeld zijn van de soorten informatie en de soorten informatiebronnen die er zijn. Daarnaast heeft elke vakgebied en elk onderwerpsdomein ook weer zijn eigen specifieke informatiesoorten en informatiebronnen, maar daar zullen we in een algemene tekst als deze uiteraard nauwelijks op in kunnen gaan. Ten behoeve van het overzicht proberen we de soorten informatie en informatiebronnen in een aantal categorieën te verdelen. Dat gebeurt hier op grond van een drietal verschillende indelingskarakteristieken:

- de aard van de informatie zelf,
- de aard van de informatiecollectie,
- de aard van de inhoud van de informatiebron.

1.1. Aard van de informatie

Een manier om informatiebronnen in categorieën in te delen, is dat te doen op basis van de aard van de informatie. Onder meer afkomst, ontstaansgeschiedenis en doel vormen daarbij belangrijke kenmerken. Hieronder vijf globale categorieën en een indicatie waarin die gepubliceerd worden en in welke vorm ze voorkomen.

1.1.1. Theorie.

Hierbij moet gedacht worden aan vastgelegde, uitgekristalliseerde kennis, opinies, inzichten, meningen en dergelijke.

Dit soort informatie is te vinden in allerlei verschillende soorten "documenten": onder meer in leerboeken, handboeken en monografieën, in (vak)encyclopedieën (ook de Wikipedia), in overzichtsartikelen (reviews) en stand-van-zaken publicaties en ook in FAQ's. FAQ's (documenten met antwoorden op "frequently asked questions") zullen meestal uitsluitend in digitale vorm beschikbaar zijn, alle andere genoemde documentsoorten kunnen zowel in papieren als in digitale vorm voorkomen.

1.1.2. Onderzoeksresultaten.

Dit betreft de resultaten van allerlei soorten onderzoek. Het kan daarbij gaan om echt proefondervindelijk onderzoek (experimenten, observaties van proefpersonen, resultaten van enquêtes en dergelijke), om theoretische berekeningen en modellen, om resultaten van computersimulaties, om beschrijvingen en rapportages van praktische toepassingen, om de neerslag van praktische ervaringen en know-how.

Dit wordt voor het grootste deel gepubliceerd in artikelen in wetenschappelijke en vaktijdschriften. Veel daarvan zijn niet alleen op papier maar ook digitaal beschikbaar, zij het meestal niet voor iedereen gratis toegankelijk. Daarnaast zijn onderzoeksresultaten te vinden in congresbijdragen, in rapporten, in proefschriften, in projectbeschrijvingen en projectverslagen, in preprints (voorpublicaties van in officiële tijdschriften te verschijnen publicaties) en soms zelfs op weblogs, discussielijsten en nieuwsgroepen.

1.1.3. Bestuurlijke informatie.

Hierbij gaat het om allerlei soorten informatie die voortvloeien uit bestuursprocessen van de overheid of die ten behoeve van bestuurlijke besluitvorming zijn verzameld of opgesteld. Gedacht kan worden aan bestemmingsplannen, streek- en structuurplannen, planologische kernbeslissingen (PKB), Milieu-effectrapportages (MER), gemeentelijke bouwvergunningen, wetgeving, kadastrale gegevens, parlementaire stukken en dergelijke. Voor een deel - wetgeving, kadastrale informatie, parlementaire informatie - wordt dit zowel in papieren als in digitale vorm op georganiseerde wijze toegankelijk gemaakt. Voor het andere deel geldt dit maar in zeer beperkte mate, het ligt vaak alleen maar fysiek ter inzage of het wordt - wel in toenemende mate - vrij ongestructureerd en op heel uiteenlopende manieren in digitale vorm door overheidsorganisaties via internet beschikbaar gesteld.

1.1.4. Gegevens, feiten, naslag.

Dit is een type informatie dat enigszins verwant is aan de eerdere categorie "theorie". Het specifieke verschil is, dat het hier gaat om concrete feiten en weetjes, zoals de prijs van een bepaald product, een telefoonnummer, het aantal medewerkers van een bepaald bedrijf, het aantal inwoners van Frankrijk in 1880, de geboortedatum van een bepaald persoon, de smeltemperatuur van een bepaalde stof. Het zijn veelal numerieke en feitelijke gegevens die alleen binnen hun specifieke context betekenis hebben.

Veel van dit soort informatie is in principe te vinden op webpagina's met specifieke product- en bedrijfsgegevens en dergelijke. Voor een belangrijk deel wordt het echter ook gepubliceerd in geaggregeerde vorm, in databases of overzichtswerken met gegevens over personen, bedrijven en organisaties, in databases of tabellenboeken met fysisch-chemische, mechanische, materiaal- en toxische eigenschappen, op websites met vergelijkende gegevens van consumentenproducten enzovoort. De beschikbare gegevens zijn meestal uit andere bronnen verzameld, vaak op juistheid of kwaliteit beoordeeld en in gestandaardiseerde vorm gepresenteerd. Daarnaast is dit soort informatie ook te vinden in statistieken, adresboeken, encyclopedieën en biografieën en ook soms in normen, octrooien, data-archieven en FAQ's.

1.1.5. Nieuws.

Hieronder verstaan we actuele, tijdgebonden berichten op allerlei terreinen: politiek, populair, sportief, financieel-economisch, bedrijfsmatig en dergelijke.

Dit nieuws wordt vooral gepubliceerd in dag- en weekbladen, als persberichten, in vakspecifieke nieuwsbladen, in attenderingsbulletins en op sommige discussielijsten en vakgerichte weblogs. Van de klassieke nieuwsbronnen als dag- en weekbladen bestaan naast de papieren versies steeds vaker tamelijk volledige digitale versies. Daarnaast bestaan er steeds meer nieuwsbronnen die uitsluitend in digitale vorm beschikbaar zijn. Wat van deze bronnen in databases wordt opgenomen, blijft meestal ook op de langere termijn - als de actualiteitsperiode voorbij is - nog beschikbaar. Wat hiervan alleen als nieuwsbericht op websites staat, kan zeer vergankelijk zijn als zo'n site geen archief heeft - elke dag is er weer vers nieuws dat voor het oude in de plaats komt.

Hoe algemeen deze vijf categorieën ook zijn omschreven, in de praktijk zul je zeker nog wel eens iets tegenkomen, dat niet precies in één hiervan is onder te brengen. Omdat het ons hier om het algemene beeld gaat, is dat gelukkig niet zo erg.

1.2. Aard van de informatiecollecties

Collecties (of bronnen) van informatie kunnen ook worden ingedeeld op basis van de mate van oorspronkelijkheid van de informatie en van de mate waarin gegevens uit diverse andere bronnen zijn verzameld en samengevat (de mate van "aggregatie" van de informatie). Men spreekt daarbij van primaire, secundaire en tertiaire informatiebronnen.

1.2.1. Primaire bronnen.

Hieronder worden die informatiebronnen verstaan waarin informatie in zijn oorspronkelijke vorm wordt gepubliceerd. Daaronder vallen veel van de bronnen die in de vorige indeling zijn gerubriceerd onder "onderzoekresultaten", onder "nieuws" en voor een deel ook wel onder "theorie". Het zijn artikelen die in kranten, tijdschriften, vakbladen en wetenschappelijke tijdschriften worden gepubliceerd. Daarnaast kan het onder meer gaan om boeken, rapporten, congresbijdragen, octrooien, standaarden, preprints, projectbeschrijvingen, biografieën, product- en bedrijfsgegevens. Artikelen verschijnen van oudsher in papieren kranten en tijdschriften. Vele daarvan hebben ook digitale versies die op de websites van hun uitgevers te vinden zijn. Sommige tijdschriften, zogenaamde E-zines, verschijnen zelfs alleen in digitale vorm op het web. Ook projectbeschrijvingen, preprints, rapporten en product- en bedrijfsgegevens zijn steeds vaker voornamelijk op de websites van individuele organisaties te vinden. Daarnaast bestaan er op het web steeds meer archieven - ook wel "repositories" genoemd - met wetenschappelijke artikelen en preprints.

The screenshot shows the homepage of 'de Volkskrant' on August 18, 2006. The main headline is 'Enkele maanden exportverbod door blauwtong' (Several months export ban due to blue tongue), with a sub-headline 'Het exportverbod voor herkauwers, ingesteld nadat de blauwtongziekte is vastgesteld onder schapen in Zuid-Limburg, zal zeker enkele maanden...'. Other news items include 'Utrecht wil fusie vier Randstadprovincies' and 'Vredesleger Libanon krijgt vorm'. The sidebar on the right features a 'QUIZ' section for 'het Witte Boekje' and 'het Groene Boekje', a 'DOSSIER' section for 'LOWLANDS 2006', and a 'WEBWINKEL' section for 'CD BOX S.JOSTAKOVITSJ'. The bottom of the page shows a 'HUIZENMARKT' section with 'TE KOOP benton' and another 'DOSSIER' section for 'GROEN OF WIT'.

Voorbeeld van primaire bron op internet: de Volkskrant

1.2.2. Secundaire bronnen.

Informatie uit primaire bronnen wordt vaak op erg versnipperde wijze aangeboden, in individuele kranten en tijdschriften, op afzonderlijke websites, in losse boeken of rapporten. Om daar toch enig overzicht over te houden, zijn zogenaamde secundaire informatiebronnen ontstaan, waarin gegevens *over* primaire informatie worden verzameld. De eerste hiervan dateren al uit het begin van de 19de eeuw. Secundaire informatiebronnen zijn vrijwel altijd gericht op een specifiek inhoudelijk thema (bijvoorbeeld "de psychologie" of "de reclame"), op een bepaald type primaire documenten (bijvoorbeeld rapporten), op een bepaald soort informatie (bijvoorbeeld bedrijfsgegevens of toxische eigenschappen) of op de collectie primaire informatie die op een bepaalde plaats aanwezig is (bijvoorbeeld een catalogus met het boekenbezit van een bepaalde bibliotheek).

Secundaire bronnen bevatten meestal niet de complete primaire informatie, maar in gestandaardiseerde vorm een extract daaruit. Elke beschrijving van een primair informatie-item heeft daarin dus dezelfde structuur. Daarbij hoort ook een verwijzing naar de vindplaats van de primaire informatie zelf. Dat is zeker het geval bij bibliografische databases (zie 1.3.1) en bij bibliotheekcatalogi (bijvoorbeeld een planknummer), maar natuurlijk ook bij onderwerpsguiden op internet (via een web-link). Verder bieden ze de gebruiker vaak toegevoegde waarde door eveneens gestandaardiseerde, niet in die vorm in de primaire informatie zelf aanwezige karakterisering van de betreffende informatie, bijvoorbeeld met trefwoorden en gestandaardiseerde formele beschrijvingselementen (bijvoorbeeld de taal van de oorspronkelijke publicatie). Tezamen worden dit wel "metadata" genoemd.



The screenshot shows the ABC Direct website interface. At the top, there is a navigation bar with links for Home, Over ABC, Adverteren, Producten, and Contact. Below this is a search bar with the text 'web design' and a dropdown menu for 'Nederland'. The search results are displayed in a table format, listing companies such as Technisch Bureau Lindhout B.V., Webnuance design & development, 21st Century Webdesign, and others. Each entry includes the company name, address, and contact options like Email and Website.

Bedrijfsnaam	Adres	Contact
Technisch Bureau Lindhout B.V.	Frans Halsstraat 44 3443 TA Woerden	Email Website
Webnuance design & development	De Veken 214 B 1716 KJ Opmeer	Email Website
21st Century Webdesign	3328 TD Dordrecht	
2morrow Visualisatie & Animatie	7534 KA Enschede	
A & A Automobiel Advies BV	2012 JL Haarlem	
A1 Boulevard Internet Service v.o.f.	4614 EE Bergen op Zoom	
ABC Communications	5632 MT Eindhoven	
Abraxax bv	2627 AT Delft	
ACB Wommels BV	8731 CD Wommels	
AccessWeb, webdesign & internet consultancy	4823 EB Breda	
ACE Colour Center Heerlen B.V.	6411 EK Heerlen	
ACE Group bv	2235 CD Valkenburg (ZH)	
Adcon IT Management B.V.	2632 DD Nootdorp	
Addesign	5616 GB Eindhoven	
Addvision B.V.	3364 DG Sliedrecht	

Voorbeeld van secundaire bron met bedrijfsgegevens: ABC Direct

Voorbeelden van dit type informatiebronnen zijn vakbibliografieën, bibliografische databases, onderwerpsguiden of webwijzers (resource guides) op internet, directories,

adresboeken, tabellenboeken en bestanden met bedrijfsgegevens of materiaal-eigenschappen. Tot deze categorie kunnen ook nog wat twijfelgevallen gerekend worden, die soms ook als primaire bronnen worden beschouwd, zoals overzichts-artikelen of het soort publicaties dat wel als "annual reviews" wordt aangeduid en ook bijvoorbeeld FAQ's en encyclopedieën. Zij bevatten geen oorspronkelijke informatie die niet al elders in primaire bronnen te vinden is, maar ze missen anderzijds vrijwel altijd de gestandaardiseerde vorm en additionele karakterisering die een belangrijk kenmerk van veel secundaire bronnen is. Bovendien worden ze meestal - net als primaire informatie - in de vorm van (wetenschappelijke) tijdschriftartikelen of boeken gepubliceerd. Daarnaast zouden ook vakgerichte weblogs (dit in tegenstelling tot de grote meerderheid aan persoonlijke weblogs) tot secundaire bronnen gerekend kunnen worden. Veelal worden daarin berichten en meningen doorgegeven, met verwijzing naar webpagina's waarop die berichtgeving of mening gefundeerd is. (Hetgeen overigens in toenemende mate zelf ook weer weblog-berichten zijn).

1.2.3. Tertiaire bronnen.

Zoals informatie uit primaire bronnen verzameld en geaggregeerd wordt in secundaire bronnen, zo zijn er ook bronnen waarin gegevens over secundaire bronnen worden verzameld. Dit worden wel tertiaire informatiebronnen genoemd.

Hieronder vallen in de eerste plaats "bibliografieën van bibliografieën" (vooral voor gedrukt materiaal) en database-guides, zoals de toonaangevende "Gale directory of online, portable and internet databases". In die laatste worden, naast gegevens over secundaire digitale bronnen, overigens ook gegevens over verzamelbronnen met primaire informatie opgenomen. Daarnaast bestaan op internet enkele overzichten van onderwerpsguides die tot dit type gerekend kunnen worden (bijvoorbeeld: Argus Clearinghouse, WWW Virtual Library en Startnederland.nl). Helaas is hun dekking vrij beperkt en wordt de eerstgenoemde niet meer geactualiseerd.

Communications and Media : en · es · fr · zh

The WWW Virtual Library
Communications and Media

Quick search:

Broadcasters	<p>Broadcasters This page contains a list of national and international TV and radio broadcasters, together with other relevant links. see also: Publishers. this resource in English is indexed under: Communications and Media.</p>
Digital Rights Management	<p>Digital Rights Management Digital Rights Management (DRM) describes a range of technologies which allow control of distribution and access to digital information, typically - but not restricted to - mass-media content (e.g. books, music, movies), software or data files (e.g. documents, spreadsheets, databases). this resource in English is indexed under: Business and Economics, Communications and Media, Computing and Computer Science, Engineering, Law.</p>
Publishers	<p>Publishers This page contains links to WWW pages relevant to publishers. see also: Broadcasters. this resource in English is indexed under: Communications and Media.</p>
Telecommunications	<p>Telecommunications The Telecoms Virtual Library is a free index of telecommunications companies and currently holds 8343 links to other telecommunications sites worldwide, indexed by subject. It is part of the World Wide Web Virtual Library and has been hosted by Analysys, one of the world's leading telecommunications consulting and research firms, since 1994. this resource in English is indexed under: Communications and Media, Engineering.</p>

See also : [Transportation](#)

Voorbeeld van tertiaire bron: WWW Virtual Library

1.3. Aard van de inhoud van de informatiebron

Voor informatie die is opgenomen in databases of andere goed doorzoekbare systemen, is ook een indeling mogelijk op basis van de aard van de daarin opgenomen informatie. De specifieke vorm en structuur van de gegevens in die databases bepalen namelijk voor een belangrijk deel de zoekmogelijkheden en de zoekmethoden die je kunt toepassen. Dit leidt tot een indeling in databases met bibliografische informatie, met full-text informatie en met feitelijke gegevens. Gezien de aard van deze indelingskarakteristiek, worden hierbij systemen uitgesloten waarin informatie toegankelijk wordt gemaakt via lijstjes hyperlinks in navigeerbare systematische indelingen. Daardoor passen niet alle in de eerdere indelingen genoemde soorten informatie (bronnen) in deze driedeling.

1.3.1. Bibliografische databases

Dit zijn databases met informatie die vrijwel altijd tot de categorie secundaire informatiebronnen behoort, maar die niet die hele categorie omvat. In bibliografische databases wordt informatie *over* diverse soorten primaire publicaties verzameld en doorzoekbaar gemaakt. Daarbij kan hetzij een specifiek inhoudelijk thema, hetzij een bepaald type primaire documenten, hetzij een bepaalde fysieke collectie als verzamelcriterium zijn genomen.

Voorbeelden waartoe de website van de mediatheek toegang geeft:

- de catalogus van de mediatheek van de Leeuwenburg,
- ERIC (artikelen en rapporten op het gebied van onderwijs en opvoedkunde),
- Marketing Data (artikelen en berichten over zakelijke en consumentenmarkten),
- Picarta (bij Nederlandse bibliotheken beschikbare boeken en tijdschriftartikelen).

The screenshot shows the Marketing Data website interface. At the top, there are navigation tabs: 'Introductie', 'Tips bij het zoeken', 'Over MD', and 'Disclaimer'. The main header area displays 'Zoekresultaten' and '1 t/m 100 van 932'. On the left side, there is a sidebar with a search box and a list of categories under 'Rubrieken', including 'Marketing & Reclame' and 'Publieke dienstverlening'. The main content area shows a list of search results. The first result is highlighted: 'Uitgeverij Free Musketeers geeft alles uit' from 'Marketingstrategie (algemeen); Uitgeverijen; Algemeen Dagblad (16-8-2006)'. Below this, there is a detailed view of the selected article, showing its title, a brief description, and metadata such as 'bronnen', 'onderwerpen', and 'landen'.

Voorbeeld van bibliografische database: Marketing Data

In dit soort databases kan meestal gezocht worden op trefwoorden, op woorden uit titels en samenvattingen (ontbreekt meestal in bibliotheekcatalogi) van de oorspronkelijke publicatie, op auteursnamen en op diverse formele kenmerken van de oorspronkelijke publicatie, zoals taal, jaar van uitgave e.d. Hoewel de databases niet de volledige inhoud van de oorspronkelijke artikelen bevatten, zijn er wel steeds meer systemen waarmee direct kan worden doorgelinkt naar elders beschikbare digitale versies van die volledige teksten (meestal alleen voor gebruikers/organisaties die daarvoor een licentie of abonnement hebben).

1.3.2. Full-text databases

Dit zijn databases met informatie die vrijwel altijd tot de categorie primaire informatiebronnen behoort, maar die zeker niet die hele categorie omvat. Het zijn databases waarin volledige publicaties zijn verzameld en doorzoekbaar gemaakt, die veelal afkomstig zijn uit meer gelijksoortige primaire bronnen. Deze primaire bronnen beperken zich meestal tot artikelen uit kranten, tijdschriften en andere periodiek verschijnende publicaties. Daarbij kunnen onder meer de volgende verzamelcriteria gehanteerd zijn: een inhoudelijk thema (bijvoorbeeld bedrijfseconomisch nieuws of biomedische wetenschappelijke artikelen), een bepaald type bronnen (bijvoorbeeld Nederlandse krantenartikelen) of de uitgaven van een bepaalde uitgever (bijvoorbeeld de inhoud van meer dan 1500 wetenschappelijke tijdschriften van Elsevier Science). Daarnaast zijn er ook full-text databases met de volledige inhoud van (vak-) encyclopedieën, leerboeken, rapporten, wetteksten en dergelijke.

Voorbeelden waartoe de website van de mediatheek toegang geeft:

- Business Source Elite (artikelen uit >1000 vaktijdschriften),
- LexisNexis Newsportal (artikelen uit landelijke en regionale dagbladen),
- SDU Wettenbank (Nederlandse wet- en regelgeving en jurisprudentie).

The screenshot shows the LexisNexis Newsportal search interface. It features a red header with the LexisNexis logo and a left sidebar with navigation links. The main search area is titled 'Nederlandstalige nieuwsbronnen' and includes a search box, a list of news sources with checkboxes, a 'zoekwoorden' section with three search input fields, and a 'Datum' section with date range selectors. At the bottom are 'Zoeken' and 'Wissen' buttons.

Voorbeeld van full-text database: LexisNexis Newsportal

Deze eerste twee voorbeelden bieden weliswaar primaire informatie, maar wel al geaggregeerd, dat wil zeggen dat het materiaal is verzameld van de uitgevers van de afzonderlijke primaire bronnen.

In dit soort systemen kan meestal alleen gezocht worden op woorden zoals die in de volledige tekst van de artikelen zelf voorkomen. In sommige gevallen ook specifiek op woorden uit titels of koppen en op auteursnamen. Slechts in uitzonderlijke gevallen ook op toegekende trefwoorden. In zekere zin kunnen ook de grote zoekmachines op internet tot deze categorie worden gerekend. Zij doorzoeken immers ook volledige teksten en bieden directe links naar de betreffende documenten (webpagina's, PDF-documenten en dergelijke). Zie hiervoor ook paragraaf 1.4.

1.3.3. Bestanden met feitelijke gegevens

Dit zijn databases waarvan de inhoud niet uit tekstdocumenten bestaat, maar uit gestructureerd opgeslagen en gepresenteerde feitelijke gegevens. Het merendeel van deze informatie zal in principe als secundair moeten worden aangemerkt. Het is in de meeste gevallen geaggregeerde, gestandaardiseerde en vaak ook geëvalueerde en op betrouwbaarheid beoordeelde informatie die is ontleend aan andere, soms heel diverse, meestal primaire bronnen. Wat voor soort bronnen dat zijn, is afhankelijk van de aard van de opgenomen gegevens. Gegevens in een bestand met bedrijfsgegevens kunnen gebaseerd zijn op bij bedrijven opgevraagde jaarverslagen, op gegevens op de websites van bedrijven of op gegevens die een bedrijf op grond van wettelijke voorschriften bij een overheidsorganisatie heeft ingediend of heeft openbaar gemaakt. Gegevens in een materiaalkundig bestand zullen vaak gebaseerd zijn op resultaten van wetenschappelijk onderzoek dat oorspronkelijk is gepubliceerd in artikelen en congresverslagen of in afzonderlijk uitgegeven tabellenboeken.

		Naam van de Organisatie	Bedrijfsopbrengsten dz EUR Laatste jaar	Aantal medewerkers Laatste jaar	Laatst beschikb. jaar	Kantoor stad	Hoofdactiviteit code (BIK)	Handelsregister nummer
1.	<input type="checkbox"/>	PCM Holding B.V.	652.639	2.901	2005	AMSTERDAM	65234	34172906
2.	<input type="checkbox"/>	PCM Uitgevers B.V.	640.552	3.034	2004	AMSTERDAM	221101	33123541
3.	<input type="checkbox"/>	PCM Algemene Boeken B.V.	94.259	463	1993	UTRECHT	221101	33012882
4.	<input type="checkbox"/>	Stichting Pensioenfonds PCM Uitgevers	24.045	n.b.	2003	ROTTERDAM	66021	34114261
5.	<input type="checkbox"/>	Bouwbedrijf P.C.M. Bakken BV	3.685	63	2004	VALKENSWAARD	452111	17070424
6.	<input type="checkbox"/>	P.C.M. Clous Beheer BV	n.b.	n.b.	2005	S-HERTOGENBOSCH	65233	16046605
7.	<input type="checkbox"/>	P.C.M. Bakken Beheer BV	n.b.	n.b.	2004	VALKENSWAARD	65234	17011559
8.	<input type="checkbox"/>	P.C.M. van Dijk Ottersum BV	n.b.	2	2001	OTTERSUM	0123	12024149
9.	<input type="checkbox"/>	P.C.M.G. Sommers Beheer BV	n.b.	1	2004	MERSELO	65234	12035597
10.	<input type="checkbox"/>	P.C.M. van Gerven Beheer B.V.	n.b.	2	2005	WEERT	65234	13034358
11.	<input type="checkbox"/>	Pattern Consultants Maastricht BV	n.b.	6	1999	GRONSVELD	748752	14061499
12.	<input type="checkbox"/>	P.C.M. van der Velden Beheer BV	n.b.	1	2004	LIEMPDE	65234	16063683
13.	<input type="checkbox"/>	P.C.M. Schellekens BV	n.b.	1	2001	OOSTERHOUT NB	74121	17040066
14.	<input type="checkbox"/>	P.C.M. de Groot Beheer B.V.	n.b.	1	1998	EINDHOVEN	65234	17046726
15.	<input type="checkbox"/>	P.C.M.C. Bosman Beheer BV	n.b.	1	2004	WAALRE	65223	17082110
16.	<input type="checkbox"/>	P.C.M. Clous Mode BV	n.b.	42	2005	S-HERTOGENBOSCH	624221	17156731
17.	<input type="checkbox"/>	PCM Projectmanagement & Support BV	n.b.	2	2004	LOON OP ZAND	748751	18070329
18.	<input type="checkbox"/>	P.C.M. Wolf Beheer BV	n.b.	1	1997	DONGEN	65234	18129584
19.	<input type="checkbox"/>	P.C.M. van den Broek Beheer BV	n.b.	1	2003	OOSTERHOUT NB	65234	20037695
20.	<input type="checkbox"/>	Advocatenpraktijk mr. P.C.M. Dirven BV	n.b.	2	2005	ETTEN-LEUR	74111	20094097
21.	<input type="checkbox"/>	P.C.M. Maas Holding BV	n.b.	1	2002	BREDA	65234	20099927
22.	<input type="checkbox"/>	PCM Consult BV	n.b.	1	2004	DORDRECHT	74141	23082094
23.	<input type="checkbox"/>	P.C.M. Peerson BV	n.b.	3	2004	ALMERE	011214	24185629
24.	<input type="checkbox"/>	P.C.M. van Mil Holding BV	n.b.	1	2004	MAASLAND	65234	24260087
25.	<input type="checkbox"/>	PCM Holding BV	n.b.	1	2004	SCHIEDAM	65234	24276862

Voorbeeld van bestand met feitelijke gegevens: Reach bedrijfsinformatie

Voorbeelden waartoe de website van de mediatheek toegang geeft:

- Euromonitor (gedeeltelijk; o.a. markt- en economische informatie per land),
- Reach (bedrijfsgegevens van Nederlandse bedrijven),
- Retriever (advertentietarieven, oplages e.d. van Nederlandse bladen).

Voor het zoeken in dit soort systemen en bij het presenteren van gevonden gegevens wordt vaak gebruik gemaakt van de sterk gestructureerde vorm van de beschikbare gegevens. Je kunt het vaak beschouwen als een soort spreadsheets.


1.4. Wel of niet op het web?

Nu het web een steeds belangrijker bron van informatie vormt, valt naast de drie genoemde soorten indelingen nog een vierde voor te stellen, namelijk een indeling in wel en niet op het web aanwezige informatie. Daaraan kleeft echter het bezwaar dat het nogal moeilijk is eenduidig te definiëren wat wordt bedoeld met "op het web aanwezig". Vrijwel alle digitale informatie is namelijk op één of andere manier wel via het web *bereikbaar*. Maar veel daarvan is opgeslagen in afzonderlijke databases die elk apart doorzocht moeten worden. Veel is ook alleen toegankelijk voor hen die persoonlijk, via hun werkgever, via hun onderwijsinstelling of via een (andere) bibliotheek een licentie hebben voor toegang en gebruik van de betreffende informatie. Dat alles maakt zo'n indeling tamelijk zinloos.


Toch is het web wel degelijk een uiterst belangrijke bron van informatie. Dat betreft dan echter al die gewone, vrij toegankelijke webpagina's en andere types documenten, die rechtstreeks via URL's en hyperlinks bereikbaar zijn. Dat is de informatie die met de grote zoekmachines als Google, Yahoo, Ask of MSN te vinden is. Die informatie vormt wel degelijk een heel speciale eigen categorie, omdat die informatie vrij toegankelijk is en omdat daarnaar met één en hetzelfde zoekhulpmiddel – namelijk die zoekmachines – gezocht kan worden. Minstens even speciaal is dat de aard van de zo gevonden informatie dwars door alle categorieën heen loopt die we hiervoor hebben onderscheiden. Van al die categorieën omvat het ook maar een tamelijk beperkt deel. Bovendien onttrekt een flink deel van de web-informatie zich aan indeling volgens die eerdere indelingscriteria. Een groot deel van deze informatie is ook niet op andere plekken en manieren te vinden. Verder heeft een belangrijk deel ervan maar een korte levensduur en is het daarna vaak definitief verdwenen.

Tot slot van deze paragraaf nog wat cijfers ter illustratie. De daarbij genoemde aantallen zijn juni 2006 gemaakte schattingen.

- Een grote zoekmachine als *Google* bevat ca. 40 miljard vrij toegankelijke webdocumenten (webpagina's, PDF's, Word-documenten, powerpoints e.d.).
- Volgens *CompletePlanet* zijn er op het web ongeveer 70.000 vrij toegankelijke gespecialiseerde zoekmachines en databases.
- Volgens de *Gale directory of online, portable and internet databases* zijn er ongeveer 25.000, grotendeels alleen tegen betaling of abonnement toegankelijke, doorzoekbare databases.
- De bij het zoekstelsel van *LexisNexis* tegen betaling toegankelijke databases bevatten ongeveer 4 miljard geselecteerde items (grotendeels full-text).
- De bij het zoekstelsel van *Dialog* tegen betaling toegankelijk databases bevatten ongeveer 3 miljard geselecteerde items (bibliografisch, full-text, en feitelijke gegevens).



CompletePlanet
The Deep Web Directory

Powered by 


HOME
HELP / FAQ
CONTACT US
ABOUT COMPLETEPLANET
TO BRIGHTPLANET

Discover over 70,000+ searchable databases and specialty search engines.

Find databases relevant to:

All of these words
Go!

ADVANCED SEARCH A comprehensive listing of [dynamic searchable databases](#). Find databases with highly relevant documents that cannot be crawled or indexed by surface web search engines.

 **All Topics >>**

<ul style="list-style-type: none"> + Agriculture + Arts & Design + Business + Computing & Internet + Education + Energy + Engineering + Environment + Family + Finance & Economics + Food & Drink 	<ul style="list-style-type: none"> + Games & Hobbies + Government + Health + Home & Garden + Humanities + Jobs & Careers + Law + Literature + Living things + Magazines & Journals + Media & Entertainment 	<ul style="list-style-type: none"> + Military + Music + News + Newspapers + People + Places + Politics + Products & Technology + Recreation + References + Regional 	<ul style="list-style-type: none"> + Religion + Science + Search Engines + Shopping + Social Sciences + Sports + Transportation + Travel + Weather
--	---	--	---

Technical White Papers: [The Deep Web](#)
[Why is Standard Search Alone Inadequate to Meet Real Business Needs?](#)

Copyright © 2000-2004, BrightPlanet Corp. All rights reserved.
 Privacy and site use policies. Problems? Report it here.

2 Informatiecollecties en hun zoekhulpmiddelen

In het vorige hoofdstuk zagen we dat de begrippen informatiesoort en informatiebron nogal dicht tegen elkaar aan zaten. Je zou kunnen zeggen dat de informatiebronnen de meer concrete systemen zijn, waarin “exemplaren” (boeken, krantenartikelen, foto’s) van één of meer informatiesoorten verzameld en aangeboden worden. In zekere zin zijn informatiebronnen dus – meestal virtuele – collecties van dergelijke exemplaren. Zulke informatiecollecties zijn natuurlijk pas nuttig als er ook een zoekhulpmiddel bij is waarmee je de daarin opgeslagen informatie ook metterdaad kunt vinden en er toegang toe kunt krijgen. Vrijwel elke informatiecollectie heeft dus ook een zoekstelsel dat daar bij hoort. Een paar voorbeelden:

- een bibliotheekcatalogus is een collectie beschrijvingen van de boeken en andere objecten in een bibliotheek of mediatheek, met daarbij een zoekstelsel om te kunnen zoeken naar boeken van een bepaalde auteur, over een bepaald onderwerp, uit een bepaald jaar en dergelijke;
- de database van de Nederlandse Vereniging van Makelaars (op Funda.nl) bevat een collectie gegevens en beschrijvingen van huizen die op dat moment te koop zijn, met daarbij een zoekstelsel waarmee je kunt selecteren en zoeken op specifieke kenmerken die voor kopers van huizen van belang zijn;
- de Newsportal van LexisNexis bevat een collectie teksten van Nederlandse krantenartikelen, met daarbij een zoekstelsel dat er vooral op gericht is om artikelen over bepaalde onderwerpen te kunnen vinden, maar waar je je bijvoorbeeld ook tot een bepaalde krant kunt beperken;
- ook een webzoekmachine als Google is een zoekhulpmiddel dat bij een bepaalde collectie hoort, namelijk de collectie van zoveel mogelijk – intussen al enkele tientallen miljarden – webpagina’s en andere op het web aanwezige documenten (PDF’s, Word-documenten, PowerPoints e.d.) die wereldwijd toegankelijk zijn.

Uit bovenstaande voorbeelden kon je al afleiden dat elk zoekhulpmiddel zijn eigen karakteristieken en zijn eigen specifieke zoekmogelijkheden heeft, die direct aansluiten bij de aard en het doel van de informatie die in de bijbehorende collectie zit. Het zoekstelsel van een bibliotheekcatalogus biedt andere mogelijkheden dan een webzoekmachine; het zoekstelsel van Funda.nl zit heel anders in elkaar dan dat van de krantendatabase in Newsportal.

Wil je goed naar informatie zoeken, dan zul je dus een duidelijk idee moeten hebben van de informatiesoort(en) waarnaar je op zoek bent, in welke informatiecollecties die te vinden zijn en hoe je de specifieke zoekhulpmiddelen gebruikt die bij die informatiecollecties horen.

Ook al kunnen er op detailniveau dus heel veel verschillen zijn tussen de diverse zoeksystemen, wat meer globaal kijkend kun je ze toch tot een paar basistypen terugbrengen:

1. Waar je bij het woord “zoekstelsel” waarschijnlijk het eerst aan denkt, zijn de systemen met zoekhokjes, waarin je zelf iets moet invullen en waarbij je dus helemaal zelf je zoekvraag zult moeten formuleren. Dat ken je namelijk van Google. Maar ook de meeste bibliotheekcatalogi en een heleboel databases, zoals de al vaker genoemde Newsportal, werken zo. Dergelijke systemen zien er

meestal vrij eenvoudig uit. Zulke invulhokjes – of dat nu bij Google is of bij de bibliotheekcatalogus – nodigen als het ware uit om iets in te vullen. Toch zul je merken dat zulke systemen in de praktijk vaak minder eenvoudig en gebruiksvriendelijk zijn dan ze er uitzien. Het blijkt vaak moeilijk om de juiste vragen te stellen, want hoe weet je welke woorden precies gebruikt zijn in de documenten (of in de beschrijvingen van de documenten) waarnaar je op zoek bent? Deze zoeksystemen zoeken immers alleen maar letterlijk naar wat je als zoekvraag intikt.

- Die problemen heb je vaak wat minder bij strakker gestructureerde databases. In de daarbij horende zoeksystemen kun je bepaalde kenmerken waarop je wilt selecteren meestal uit keuzelijstjes kiezen, zodat je niets zelf hoeft te bedenken. Veel van de in paragraaf 1.3 genoemde bestanden met feitelijke gegevens bieden dergelijke mogelijkheden. In het ook al genoemde systeem van Funda.nl kun je bijvoorbeeld uit lijstjes kiezen wat voor type woning je zoekt, in welke prijsklasse, met of zonder tuin, enzovoort.

The screenshot shows the Funda.nl search interface with the following elements:

- Zoek op plaats:** A dropdown menu set to 'Noord-Holland'. Below it, instructions: 'Kies max. 5 plaatsen: 1. Kies een provincie, 2. Kies dan een plaats, 3. Gebruik pijlen voor toevoegen/verwijderen'. A list of locations includes 'Oude Nierdorp', 'Oudendijk Nh', 'Ouderkerk Aan De Amst', 'Oudeschild', and 'Oudesluis'. To the right, 'Geselecteerde plaats(en):' shows 'Amstelveen' and 'Amsterdam'.
- Prijsklasse:** Two dropdown menus for price range, set to '€ 200.000' and '€ 350.000'.
- Optionele zoekcriteria:** A section with various filters:
 - Soort bouw:** 'Bestaande bouw'
 - Soort object:** 'Woonhuis'
 - Soort woning:** 'Geen voorkeur'
 - Type woning:** 'Eengezinswoning' (highlighted)
 - Aantal kamers:** 'Eenvoudige woning', 'Herenhuis', 'Villa', 'Landhuis'
 - Woonoppervlakte:** 'Bungalow', 'Woonboerderij', 'Grachtenpand', 'Woonboot'
 - Perceeloppervlakte:** 'Stacaravan', 'Landgoed'
 - Ligging:** 'Aan park'
 - Toon aanbod sinds:** '10 dagen'
 - Bijzonderheden:** A list of checkboxes: 'Tuin' (checked), 'Balkon', 'Garage', 'Berging', 'CV' (checked), 'Openhaard', 'Zwembad', 'Airco'.

Voorbeeld van Funda-zoekscherm

- Ook voor nauwelijks gestructureerde informatie bestaan systemen waarin je niet zelf zoekwoorden hoeft te bedenken, omdat de informatie daarin al systematisch op onderwerp is ingedeeld. Dergelijke systemen kom je ook op internet veelvuldig tegen in de vorm van zogenaamde onderwerpsgidsen. De systematische onderwerpsindeling is meestal een soort hiërarchische boomstructuur, die zich steeds verder vertakt in steeds specifiekere deelonderwerpen. Op internet kom je algemene gidsen tegen, waarin in principe alle onderwerpen terug te vinden zijn, zoals de Yahoo-directory, de OpenDirectory of (wat simplistischer en nauwelijks

hiërarchisch) Startpagina.nl. Daarnaast zijn er ook talloze gespecialiseerde gidsen, die zich tot een bepaald onderwerpsgebied beperken, zoals de in paragraaf 1.2 bij secundaire bronnen genoemde onderwerpsgidsen; denk hierbij ook aan de (overigens ook niet erg hiërarchische) Startpagina-dochters. Voordeel van veel dergelijke systemen is dat je steeds specifiekere onderwerpscategorieën kunt kiezen en aanklikken, zodat je de gewenste informatie in feite vindt door browsen of navigeren door de boomstructuur van de onderwerpsindeling. Dergelijke systemen zijn dus vooral geschikt in situaties waarin je zelf moeilijk actief kunt omschrijven waarnaar je op zoek bent, maar waarin je wel (passief) in keuzemenu's herkent onder welk onderwerpsgebied de gezochte informatie waarschijnlijk te vinden is.

The screenshot shows the Yahoo! News Directory interface for 'Internet Broadcasts'. At the top, there is a search bar with options to search 'the Web', 'the Directory', or 'this category'. Below the search bar, the page title is 'Internet Broadcasts' with links for 'Email this page', 'Suggest a Site', and 'Advanced Search'. The breadcrumb trail is 'Directory > News and Media > Internet Broadcasts'. There are also links for 'Save to My Web' and 'What's This?'. The main content is divided into 'CATEGORIES' and 'SITE LISTINGS'. Under 'CATEGORIES', there are 'Top Categories' and 'Additional Categories'. 'Top Categories' includes 'News (30)', 'Podcasting and Audioblogging (99)', 'Video Blogs (Vlogs) (32)', and 'Video Sharing (39) NEW'. 'Additional Categories' includes 'College and University (94)', 'Computers and Internet@', 'Content Aggregators@', 'Entertainment@', 'Finance and Investment@', 'Health@', 'Multimedia Technology@', 'Music@', 'Regional (22)', 'Science@', 'Services and Software@', 'Sports@', 'Talk Format (12)', and 'Web Directories (16)'. 'SITE LISTINGS' are sorted by 'Popularity' and show 'RAIN: Radio And Internet Newsletter', 'Kill Radio', and 'DTV'. On the right side, there are 'SPONSOR RESULTS' for 'Bill Campbell Ringtone', 'Toddler Rain Coats, Toddler Rain Boots', 'The Rain', 'Waterproof Breathable Rain Jackets', and 'NoCost Ringtones - Rain'.

Voorbeeld van een stukje uit de boomstructuur van de Yahoo-directory

In het vervolg van deze tekst zal vooral aandacht worden besteed aan de eerste hier genoemde categorie zoeksystemen. Daarbij moet (en kan) je namelijk zelf het meest sturing geven aan de manier waarop je je zoekacties aanpakt.

3 Zoekstrategie en zoekmethode

Uit de voorgaande hoofdstukken is al duidelijk geworden dat, voor een degelijk opgezette zoektocht naar informatie, kennis nodig is van informatiesoorten die voor bepaalde informatiebehoefte zinvol zijn, dat je moet weten in welke bronnen of informatiecollecties dat soort informatie te vinden is en dat je de mogelijkheden en kenmerken van de daarbij horende zoekhulpmiddelen moet kennen. Die kennis zorgt dat je de juiste keuzes kunt maken, hoe je een zoektocht gaat aanpakken. Toch vormen die keuzes nog maar de eerste stappen. Van minstens evenveel belang is hoe je uiteindelijk je zoekacties concreet gaat aanpakken in die afzonderlijke zoekhulpmiddelen. Over de daarbij te gebruiken methodes hebben we het tot dusverre nog niet gehad. Dat zal daarom in dit hoofdstuk aan de orde komen.

Het hele traject van bepaling van benodigde informatiesoorten, van keuze van te gebruiken informatiebronnen en zoekhulpmiddelen, tot en met de concrete uitvoering van zoekacties met die zoekhulpmiddelen, wordt wel met de term zoekstrategie aangeduid. Om die zoekstrategie goed uit te voeren is in feite een heel plan van aanpak nodig. Op dat "zoekplan" komen we in een volgend hoofdstuk nog nader terug, nadat we eerst wat meer idee hebben gekregen van de zoekmethodes die je in allerlei systemen kunt toepassen.

3.1 De zoekvraag

Uitgangspunt van elke zoekactie is uiteraard de informatievraag. En informatievragen heb je in allerlei soorten en maten. Een belangrijke karakteristiek is of het je gaat om een echte onderwerpsvraag of om een zogenaamde "known item" vraag. Met dat laatste wordt bedoeld dat je bijvoorbeeld weet dat een bepaald boek bestaat, en dat je nu op zoek bent of het in een bepaalde bibliotheekcollectie aanwezig is. Een ander voorbeeld hiervan zijn de meeste vragen naar feitelijke gegevens. In welk jaar werd Mozart geboren? Bij welke temperatuur smelt ijs? Hoe ziet de vlag van Nicaragua er uit? Je weet dat het gegeven bekend moet zijn, alleen ken jij het (nog) niet. Je bent dan al tevreden met een getal, een naam of een plaatje dat het antwoord op je vraag vormt. Bij een onderwerpsvraag daarentegen ben je meestal op zoek naar teksten over een onderwerp, waarin uitleg, meningen en - toch ook wel - allerlei gegevens daarover te vinden zijn. Die teksten geven meestal aanleiding tot bestudering, interpretatie, vergelijking en beoordeling. In dit hoofdstuk gaan we vooral op dergelijke onderwerpsvragen in.

Een belangrijke karakteristiek van onderwerpsvragen is verder nog hoe diepgaand informatieonderzoek moet worden uitgevoerd. Zo zijn er situaties waarin een snelle oppervlakkige zoekactie volstaat, bijvoorbeeld omdat je je alleen even wilt oriënteren over een bepaald onderwerp. Je bent dan in feite al tevreden, zodra je maar iets hebt gevonden waaruit wat informatie over het onderwerp te destilleren valt. Men spreekt daarbij wel van een "quick & dirty" zoekactie. Daarvoor hoeft je in de meeste gevallen geen heel ingewikkelde zoekstrategieën toe te passen - al moet je uiteraard wel enige informatie over je onderwerp kunnen vinden. Er zijn echter ook situaties waarin veel diepgaander onderzoek nodig is. Waar je van een bepaald onderwerp allerlei invalshoeken moet achterhalen, of de diverse opvattingen die over een bepaald thema

bestaan. Of waar je zo volledig mogelijk alles moet vinden over een bepaald onderwerp om te zien of er nog onbekende aspecten zijn, waar je zelf onderzoek naar zou kunnen doen. Om in die gevallen uiteindelijk de gewenste informatie op te sporen, zijn meestal veel uitgekinder zoekmethodes nodig, zoals in de volgende paragrafen aan de orde komen.

Een eerste stap bij een onderwerpsvraag is een afbakening van je vraag. Daarvoor is vaak wel al enige kennis van het onderwerp nodig. Soms zul je die al hebben. In andere gevallen is daarvoor al zo'n oriënterend (quick & dirty) onderzoek nodig. Daaruit kun je dan leren wat er precies speelt, hoe breed of hoe beperkt het onderwerp is, welke aspecten en invalshoeken er aan het onderwerp zitten en welke deelonderwerpen te onderscheiden zijn. Soms ook heb je te maken met een opdrachtgever die al duidelijke ideeën heeft wat hij precies wil weten.

Stel dat je onderzoek wilt doen naar RSS (Rich Site Summary), een methode waarmee nieuwe informatie via internet gedistribueerd kan worden, zodat gebruikers die automatisch, overzichtelijk en gecombineerd uit allerlei verschillende bronnen op hun scherm kunnen krijgen. Gaat het je dan om de technische kanten daarvan of om marketingaspecten? Gaat het je om het nut voor gebruikers of om het belang voor informatieproducenten om zo hun informatie te verspreiden? Gaat het om een overzicht welke soorten informatie je zo kunt binnenhalen of welke diensten al op basis van RSS zijn ontwikkeld? Na zo'n afbakening kan die RSS-vraag uiteindelijk bijvoorbeeld zijn komen te luiden: "technieken die informatieproducenten kunnen gebruiken om met behulp van RSS een laagdrempelige informatiedienst voor specifieke doelgroepen op te zetten". Vrijwel elke informatievraag zal zo nader gespecificeerd en afgebakend kunnen (en ook moeten) worden.

3.2 Zoektechnieken

Voordat we op de verschillende praktische aanpakken van zoekvragen ingaan, zullen we eerst eens kijken naar de technieken die in allerlei zoeksystemen kunnen worden toegepast om informatie te zoeken. Daarbij beperken we ons in deze paragraaf tot de al eerder genoemde "systemen met zoekhokjes".

3.2.1 Best-match zoeken

De bij vrijwel iedereen meest bekende zoeksystemen zijn op dit moment de zoekmachines op internet, en dan vooral Google. De techniek die door die systemen wordt toegepast, wordt wel aangeduid met de term "best match" zoeken. Hierin worden technieken toegepast waarmee het systeem probeert te bepalen welke resultaten "het meest overeenkomen" met de gestelde zoekvraag. Dat houdt in dat zoekresultaten ook in een bepaalde volgorde gepresenteerd worden. Omdat die mate van overeenkomst met de zoekvraag als een maat voor de relevantie van de gevonden informatie wordt beschouwd, wordt dit wel de relevantie-volgorde of "relevance ranking" genoemd.

Wanneer je bij zoekmachines als Google maar een enkel woord als zoekvraag intikt, is het voor dergelijke systemen natuurlijk heel moeilijk om zonder verdere context te bepalen wat voor jouw vraag het meest relevant is. Daarom wordt in zulke gevallen eigenlijk alleen gekeken of jouw zoekwoord in de titel van een webpagina voorkomt, of wellicht in koppen in de tekst. Zo ja dan hebben die pagina's een grotere kans voor

jou relevant te zijn, dan een pagina waar jouw zoekwoord onopvallend ergens onder aan de pagina staat. Daarnaast wordt ook het vermoedelijk belang van de op jouw zoekwoord gevonden pagina's afgewogen op grond van het aantal andere webpagina's dat daarnaar verwijst, dat een hyperlink daarheen bevat. Hoe die afwegingen precies geprogrammeerd zijn is het geheim van elke zoekmachine. Elke zoekmachine doet dat dan ook een beetje anders.

Zoals we al zagen zul je zoekvragen vaak veel preciezer moeten afbakenen, waarvoor meer zoekwoorden nodig zijn. Als je inderdaad meer woorden intikt, wordt het voor een zoekmachine al makkelijker om de relevantie van gevonden pagina's ook nog op wat meer inhoudelijke gronden in te schatten. Er wordt dan van uitgegaan dat de woorden die een gebruiker als zoekvraag intikt, een soort zinnetje vormen. Hoe preciezer een webpagina dat zinnetje bevat, hoe beter die waarschijnlijk met de zoekvraag overeenkomt en hoe relevanter die dus waarschijnlijk zal zijn. Onderlinge afstand en volgorde van de zoekwoorden in de gevonden webpagina's speelt daar dus een belangrijke rol.

Als je meer woorden als zoekvraag intikt, vertoont een pagina die niet al die woorden bevat, maar bijvoorbeeld maar drie van de vier, ook nog wel enige "overeenkomst" met de zoekvraag. Toch leveren de meeste grote zoekmachines standaard alleen resultaten op waar de door de gebruiker gevraagde woorden allemaal in voorkomen. De omvang van het web is namelijk zo groot dat zelfs een zoekvraag op vijf woorden meestal nog zoveel resultaten met alle vijf die woorden oplevert, dat er weinig aanleiding is ook nog te laten zien wat op één na al die woorden bevat. Dat betekent dat de meeste zoekmachines standaard staan ingesteld te zoeken op "all of the words". Voor heel specialistische onderwerpen en bij zoekvragen die met heel veel zoekwoorden gespecificeerd zijn, kun je in de "advanced" versies van zoekschermen meestal ook specificeren dat het systeem moet zoeken op "any of the words" of "at least one of the words". Ook in dat geval krijgen webpagina's waarin wel alle zoekwoorden voorkomen nog altijd voorrang.

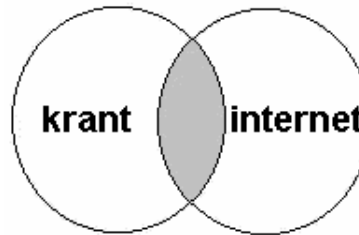
Google-advanced: je kunt kiezen of je resultaten wilt die voldoen aan "all of the words" of aan "at least one of the words".

3.2.2 Booleaans zoeken

De hiervoor besproken "best match" zoekmethode houdt er weinig rekening mee dat je voor sommige zoekwoorden wel eens synoniemen of alternatieven in je zoekactie zou willen verwerken. In zo'n geval kan het je niet schelen welk van beide woorden in het zoekresultaat voorkomt. Als je in een standaard zoekscherm van een zoekmachine twee synoniemen in je zoekvraag verwerkt, zul je namelijk alleen resultaten krijgen waarin die twee synoniemen samen voorkomen. En ook als je met "any of the words" zoekt, kun je niet aangeven welke woorden uit je zoekvraag nu precies de synoniemen zijn, waar er eigenlijk maar eentje van hoeft voor te komen. Dat soort specificaties en nog veel meer zijn wel mogelijk met de zogenaamde Booleaanse zoekmethode. In vrijwel alle zoeksystemen voor databases en bibliotheekcatalogi kan die worden toegepast. Maar veel zoekmachines op internet, zoals onder meer Google, Yahoo, AltaVista, MSN, Exalead en Ilse, blijken gebruik van die methode ook wel degelijk te ondersteunen. Bij de Booleaanse zoekmethode zijn standaard drie operatoren beschikbaar om zoekwoorden met elkaar te combineren, AND, OR en NOT.

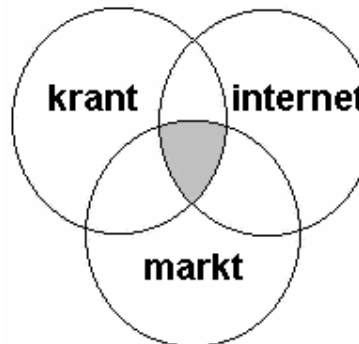
Met de AND operator geef je aan dat beide zoekwoorden in de te vinden documenten moeten voorkomen. Met bijvoorbeeld

krant AND internet



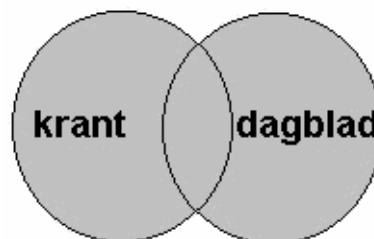
vind je dus alleen documenten waarin die twee woorden samen voorkomen. Een AND-operator zorgt dus voor een sterke inperking van je zoekresultaat. En hoe meer woorden je met AND combineert hoe strengere eisen je aan de te vinden documenten stelt, dus hoe minder je overhoudt.

krant AND internet AND markt



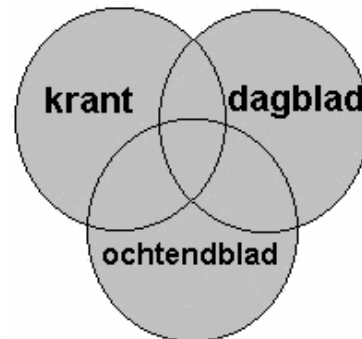
Met de OR-operator geef je aan dat maar één van beide woorden in een document hoeft voor te komen (maar beide mag natuurlijk ook). Met bijvoorbeeld

krant OR dagblad



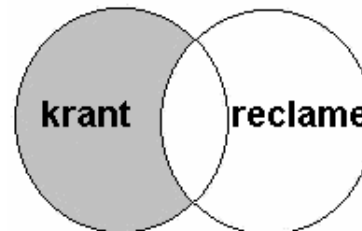
vind je dus alle documenten waarin het woord "krant" voorkomt, plus alle documenten waarin het woord "dagblad" voorkomt (en dus ook die waarin ze allebei voorkomen). Een OR-operator zorgt dus voor uitbreiding van je zoekresultaat. En hoe meer woorden je met OR combineert hoe meer je je zoekresultaat uitbreidt.

krant OR dagblad OR ochtendblad



Met de NOT-operator geef je aan dat een bepaald woord beslist niet in je zoekresultaten mag voorkomen. Met bijvoorbeeld

krant NOT reclame



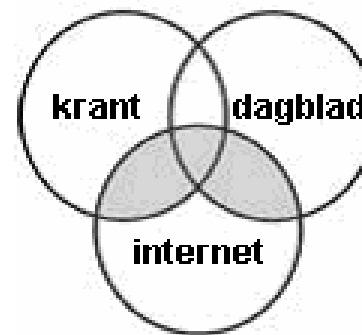
zal geen enkel document meer in je zoekresultaat zitten waarin het woord "reclame" voorkomt, ook al staat het woord "krant" er wel degelijk in. Dat is dus weer een manier om je zoekresultaat in te perken.

NB1: bij sommige zoeksystemen moet je in plaats van NOT de operator ANDNOT gebruiken ("*krant ANDNOT reclame*").

NB2: bij Google moet je in plaats van NOT een minteken direct voor het ongewenste woord zetten ("*krant -reclame*")

In een enkele zoekopdracht kun je ook verschillende operatoren combineren. Gebruik daarbij wel haakjes om aan te geven welke zoekwoorden het eerst gecombineerd moeten worden. Zonder die haakjes is het niet altijd duidelijk hoe een zoekstelsel je zoekvraag interpreteert. Dus bijvoorbeeld

(krant OR dagblad) AND internet



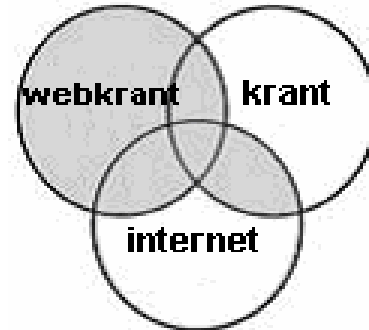
om documenten te krijgen waarin het woord "internet" voorkomt samen met het woord "krant" of het woord "dagblad". Met een andere volgorde

internet AND (dagblad OR krant)

krijg je uiteraard hetzelfde resultaat.

Met

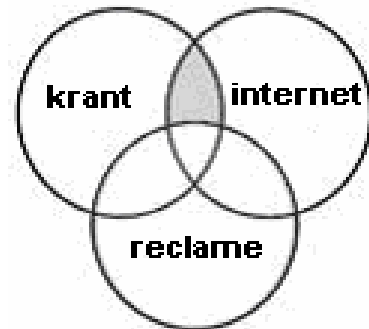
(krant AND internet) OR webkrant



breid je het resultaat van "krant AND internet" uit met alle documenten waarin alleen maar het woord "webkrant" hoeft voor te komen.

Met

(krant AND internet) NOT reclame



verwijder je juist alle documenten waarin het woord "reclame" voorkomt.

Met die Booleaanse methode kun je dus heel precies zoekvragen formuleren. Ook als je voor meer van je oorspronkelijke zoekwoorden synoniemen wilt toevoegen, is dat mogelijk. Zo zou je uiteindelijk bijvoorbeeld de volgende zoekvraag kunnen krijgen:

(krant OR dagblad OR ochtendblad) AND (internet OR web) AND markt NOT reclame

De methoden uit bovenstaande voorbeelden kunnen ook gebruikt worden bij webzoekmachines waarmee je Booleaans kunt zoeken. Je hoeft daarvoor niets extra's te doen, behalve dan dat het verplicht is de operatoren in hoofdletters in te tikken.

Nadeel van de klassieke Booleaanse zoekmethode is dat het nogal een "zwart-wit" methode is. Een document voldoet wel of niet aan je combinatie van zoektermen. Nuancering wordt daar in principe niet in aangebracht. De presentatievolgorde van zoekresultaten uit databases en bibliotheekcatalogi wordt daarom meestal alleen bepaald door het tijdstip dat de gevonden gegevens aan het systeem zijn toegevoegd, de meest recente gegevens het eerst. Dat laatste geldt overigens niet voor webzoekmachines. Ook op de resultaten van Booleaanse combinaties proberen die toch nog wat van hun methoden van relevance ranking toe te passen.

Hoe nuttig en noodzakelijk de AND-operator voor het doen van zoekacties ook is, er kleeft ook een nadeel aan. Dat twee woorden samen in een document voorkomen, betekent nog niet dat ze inhoudelijk ook echt iets met elkaar te maken hebben. Toch verwacht je dat meestal wel als je op zo'n combinatie zoekt. Zeker in gevonden documenten met veel tekst kunnen twee met AND gekoppelde woorden ver bij elkaar vandaan staan en niets met elkaar te maken hebben. Om dan toch wat preciezer te zoeken, kun je in sommige zoeksystemen aangeven dat de gezochte woorden bij elkaar in de buurt moeten staan. Bij sommige systemen (ook bijvoorbeeld de niet zo bekende webzoekmachine Exalead) kun je daarvoor de NEAR-operator gebruiken.

krant NEAR internet

betekent dan dat de beide woorden niet alleen samen moeten voorkomen, maar dat ze ook op niet meer dan 10 woorden afstand van elkaar mogen staan. Er is dan een veel grotere kans dat er een inhoudelijke relatie tussen die woorden bestaat. In andere systemen, zoals bijvoorbeeld LexisNexis Newsportal, kun je zelfs nog preciezer de maximale afstand aangeven.

krant W/3 internet

betekent daar dat er niet meer dan drie andere woorden tussen de beide zoekwoorden mogen staan. Deze operatoren zijn dus nog "strenger" dan AND en perken je zoekresultaat nog sterker in.

In de tot nu toe gegeven voorbeelden werden alle te combineren termen direct in één keer in een soms lange uitdrukking ingetikt. Niet in alle zoeksystemen is dat nodig. Er zijn ook systemen waar je je zoekactie in stapjes kunt opbouwen. Daarin krijgt dan elk tussenresultaat - meestal een resultaat-set genoemd - een volgnummer, waarna je die tussenresultaten alsnog met Booleaanse operatoren kunt combineren.

Het eerdere complexe voorbeeld zou er dan ongeveer als volgt kunnen uitzien:

- #1: *krant OR dagblad OR ochtendblad***
- #2: *internet OR web***
- #3: *markt***
- #4: *reclame***
- #5: *(#1 AND #2 AND #3) NOT #4***

3.2.3 Truncatie en woordstammen

In de voorgaande voorbeelden is steeds alleen het enkelvoud van de zoekwoorden gebruikt. Je mag er echter beslist niet van uitgaan dat elk zoekstelsel dan ook meteen documenten vindt waarin de betreffende woorden toevallig alleen in het meervoud voorkomen. De voorbeeld-zoekvragen in de vorige paragraaf zullen in het algemeen dus nog zeer onvolledige resultaten opleveren. Het is natuurlijk onhandig om voor volledige zoekresultaten zelf alle verschillende woordvormen met OR te moeten combineren. Daarvoor bestaan dan ook verschillende oplossingen.

Veel zoeksystemen bieden de mogelijkheid om aan te geven dat je wilt zoeken op alle woorden die met een bepaald woordgedeelte beginnen, zogenaamde truncatie of in

computerjargon ook wel "zoeken met wildcards". Daarmee kun je niet alleen problemen met enkel- en meervoud voorkomen, maar zoek je ook op andere woorduitgangen van je zoekwoord, zoals eventueel daarvan afgeleide werkwoordsvormen, bijvoeglijke naamwoorden of bijwoorden en vervoegingen en verbuigingen daarvan, alsook op samenstellingen met andere woorden. Een veel gebruikt symbool voor truncatie (de wildcard) is de asterix *. Met

*krant**

zoek je dan dus ook op "kranten", "krantjes", "krantenknipsel", "krantenkop", "krantenpapier", "krantenwijk", enz.

In sommige zoeksystemen kunnen ook binnenin een woord één of meer letters "gemaskeerd" worden, om op spellingsvariaties te kunnen zoeken. Met bijvoorbeeld

publi?aties

vind je dan beide schrijfwijzen van het woord, met een "c" of met een "k" (en in principe natuurlijk ook met andere letters op de plaats van het vraagteken, maar die zullen in dit geval niet voorkomen).

Helaas bieden de bekende webzoekmachines niet dit soort mogelijkheden tot truncatie. Daar zul je dus zelf alle woordvarianten, met OR gecombineerd, in je zoekvraag moeten meenemen. Het nog wat onbekende Exalead vormt op dit moment de enige uitzondering daarop.

Met truncatie kun je als zoeker zelf bepalen wat de "woordstam" is waarop je wilt zoeken, dus welk deel van het woord je wilt afkappen en door een truncatie-teken vervangen. Er zijn ook systemen die dit automatisch proberen te doen. De eenvoudigste variant hiervan is, dat bij een in enkelvoud ingetikt zoekwoord ook automatisch op het meervoud wordt gezocht en omgekeerd. Een stapje verder gaan systemen die elk zoekwoord tot hun taalkundige woordstam reduceren, door standaard achtervoegsels van die woorden te verwijderen. Voor allebei die methoden heeft het zoekstelsel dus enige kennis nodig van de taal waarin gezocht wordt; het moet weten wat in een bepaalde taal de regels zijn voor meervoudsvormen of welke standaard uitgangen in die taal achter een woordstam kunnen voorkomen. Systemen die automatisch op woordstammen zoeken passen in de praktijk vrijwel alleen Engelse taalregels toe. Bij de grote aanbieder LexisNexis (van Newsportal) wordt zo automatisch op Engels enkel- en meervoud gezocht en zelfs op verschil tussen Engelse en Amerikaanse spelling: *bicycle* vindt ook *bicycles*, *behaviour* vindt ook *behavior* en omgekeerd.

3.2.4 Zoekingangen; zoeken in velden

Bij veel zoeksystemen zul je standaard in alle digitaal beschikbare tekst zoeken. Bij webzoekmachines is dat de volledige tekst van de webpagina's en de PDF's en ook van de Word-documenten en Powerpoints die daarmee eventueel doorzoekbaar zijn. In dit soort documenten valt ook weinig verder te specificeren omdat er geen standaard inhoudelijke structuur in zit. Bij gegevens in bibliotheekcatalogi en andere databases ligt dat heel anders. Daarin zijn gegevens veel gestructureerder opgeslagen en van die structuur in zogenaamde velden kun je bij het zoeken gebruik maken.

Gegevens in bibliotheekcatalogi en in bibliografische databases hebben een min of meer vergelijkbare structuur waarin de oorspronkelijke publicaties worden beschreven. Dat zijn de in paragraaf 1.2.2 genoemde metadata. Hoewel het om veel meer gegevens gaat dan alleen de titels van die publicaties, worden dit ook vaak "titelbeschrijvingen" genoemd. In die titelbeschrijvingen vind je in elk geval altijd velden voor (uiteraard) de titel, voor de namen van de auteurs en voor het jaar van publicatie. Afhankelijk of het om een boek, een tijdschriftartikel of nog een ander soort document gaat, kunnen er ook velden zijn voor de uitgever van het boek, voor het aantal pagina's daarvan, voor de naam van het tijdschrift waarin het artikel is verschenen en voor de datum waarop, het afleveringsnummer waarin en de pagina-nummers waar het artikel stond. Daarnaast is er vaak ook nog een veld voor trefwoorden die aangeven waarover de publicatie gaat, voor rubrieken of categorieën op basis waarvan de collectie systematisch is ingedeeld en voor een samenvatting van de inhoud van de publicatie. Veel van deze velden kunnen als "zoekingang" worden gebruikt. Dat betekent dat je bij het zoeken kunt aangeven dat je specifiek alleen in de inhoud van een bepaald veld wilt zoeken.

Zoeken in auteursveld

Bij auteursnamen is de achternaam de primaire zoekterm om op te zoeken. De voornaam of voorletter komt daar pas achter. Bij achternamen met voorvoegsels is niet altijd duidelijk wat daarmee gebeurt. In Nederland worden die meestal niet als onderdeel van de achternaam beschouwd (zoals Piet de Vries in het telefoonboek ook onder de V gezocht moet worden). In Engels- of Franstalige systemen en ook in België worden voorvoegsels meestal wel bij de achternaam gerekend en daar vaak ook aan vast geschreven. Het is in zulke gevallen dus altijd zaak verschillende mogelijkheden uit te proberen (Vries, P. de / De Vries, P. / Devries, P.).

The screenshot shows the 'Browse Indexes' interface for LISA. At the top, there are navigation tabs: 'Combine Searches', 'Alerts', 'History', 'Command Search', 'Thesaurus', and 'Indexes'. The 'Indexes' tab is selected. Below the tabs, there is a 'Select Database & Index' dropdown menu showing '- Author Index' and a 'Change Databases' link. A search box contains the text 'sieverts' and a 'Go' button. Below the search box is a 'Browse the Index:' section with an alphabetical list of letters from A to Z. The main content area is titled 'LISA: Library and Information Science Abstracts: Author Index'. On the left, there is a section 'New Search Using Marked Terms' with two radio buttons: 'Use AND to narrow' (unselected) and 'Use OR to broaden' (selected). Below this is a green 'Search' button. On the right, there is a list of search suggestions, each with a checkbox: 'sievert m e', 'sievert mary ellen', 'sievert mary ellen c', 'sievert maryellen', 'sievert maryellen c', 'sieverts e', 'sieverts e g', 'sieverts eric', 'sieverts eric g', 'sieving pamela c', 'siew chee kheong', and 'siew k y'.

In databases worden auteursnamen meestal niet gestandaardiseerd; alfabetisch lijstje van in dit systeem voorkomende auteursnamen.

Of je voor de voornaam alleen voorletters of hele voornamen moet gebruiken, kan per systeem heel verschillend zijn. Bij bibliotheekcatalogi wordt er vaak naar gestreefd om controle uit te oefenen op de namen van (bekende) auteurs, zodat verschillende naamvormen en pseudoniemen van een zelfde auteur als zodanig herkend worden (bijvoorbeeld G.K. van het Reve = Gerard Reve = Simon van het Reve) en het dus ook niet uitmaakt op welke vorm je zoekt. Anderzijds kan daar onderscheid gemaakt worden tussen verschillende auteurs die toevallig exact dezelfde naam hebben. Bij bibliografische databases vindt dat soort controle vrijwel nooit plaats, zodat je als zoeker zelf moet gokken hoe je alle publicaties van Eric Sieverts bij elkaar moet krijgen (E. / E.G. / Eric / Eric G. / ...?).

Zoeken op titel

Bij bibliotheekcatalogi kun je meestal kiezen of je wilt zoeken op "de titel" of op "titelwoorden". In het eerste geval moet je de precieze titel van een publicatie weten, of op zijn minst het begin daarvan. Een eventueel lidwoord aan het begin van een titel moet hierbij worden weggelaten. Deze methode heeft uiteraard alleen zin voor een zogenaamde "known item search". Veel handiger is het vaak om te zoeken op "titelwoorden". Daarvoor hoef je alleen één of meer losse woorden in te tikken. Voor een "known item search" moeten dat natuurlijk een paar zeer karakteristieke woorden uit de (bekende) titel zijn. Voor een onderwerpsvraag gebruik je gewoon de zoekwoorden die je onderwerp omschrijven. Een reden om hierbij specifiek op titelwoorden te zoeken, is dat zo gevonden publicaties een veel grotere kans hebben werkelijk over het gezochte onderwerp te gaan.

Universiteitsbibliotheek Utrecht Catalogus

UUHome > Bibliotheek > Catalogus UBU

Zoeken Uitgebreid zoeken Zoekgeschiedenis De laatste titellijst Inloggen/Boeken verlengen Bewaarset Reset Help Stel uw vraag

Zoeken met losse woorden:

Kies een veld: Titelwoorden i

Vul losse woord(en) in: i Zoeken

Tik een of meer titelwoorden in
plato* mathem*
wiskun* or mathemat*
wom!n

Zoeken in een alfabetische lijst:

Kies een lijst: Titel begint met i

Vul beginwoord(en) in: i Zoeken

Tik (het begin van) een titel in
zonder het beginlidwoord
U zoekt: De hond.. / A tale.. / L'ange..
U tikt in: hond.. / tale.. / ange..

Zoeken op titelwoorden of op hele titels ("Titel begint met")

Hoewel webdocumenten niet echt in velden gestructureerd zijn, is het met zoekmachines toch mogelijk om - als enige specifieke zoekingang - op titelwoorden te zoeken (zie bijvoorbeeld het in paragraaf 3.2.1 afgebeelde Google-schermbild). Bedenk daarbij wel dat het begrip "titel" voor webpagina's een veel minder formeel gegeven is als bij boeken of tijdschriftartikelen. Het is het stukje tekst dat toevallig een HTML-code <title> heeft meegekregen.

Zoeken op trefwoorden

De meeste catalogi en databases hebben een speciaal veld voor trefwoorden. Voordeel van het gebruik van trefwoorden bij het zoeken, is dat je redelijk zeker weet dat daarmee gevonden documenten ook echt over het onderwerp gaan, waarnaar je op zoek bent. Een niet onbelangrijk nadeel is dat de manier waarop trefwoorden worden toegekend, en dus ook de manier waarmee je daarmee moet zoeken, per systeem nogal kan verschillen. In sommige systemen worden enkelvoudige woorden of begrippen gebruikt, die je zelf bij het zoeken met een AND-operator met elkaar zult moeten combineren

literatuur AND frankrijk AND twintigste eeuw

In andere systemen wordt gebruik gemaakt van zogenaamde gelede trefwoorden waarin begrippen tevoren al in één term gecombineerd zijn;

“literatuur; frankrijk; 20ste eeuw”

voor een publicatie die over 20ste eeuwse Franse literatuur gaat.

In sommige systemen zijn de trefwoorden ontleend aan een standaard lijst met toegestane trefwoorden, zodat je als zoeker moet weten op welke woorden je wel en niet moet zoeken. In andere systemen worden trefwoorden veel meer willekeurig toegekend. Wel wordt dan meestal besloten om daarbij alleen meervoudsvormen van woorden te gebruiken (of juist alleen enkelvoud).

Als je bij het zoeken gebruik wilt maken van trefwoorden, zul je dus moeten weten hoe daarmee in de door jou gebruikte catalogus of database precies wordt omgegaan. Het is dan ook niet erg zinvol om daar in deze paragraaf in nog meer detail op in te gaan.

Zoeken op systematiek

Bij veel bibliotheekcatalogi en ook bij sommige databases wordt gebruik gemaakt van een zogenaamde systematische indeling van het materiaal. Dat wil zeggen dat er sprake is van een aantal hoofdrubrieken of categorieën, in de praktijk meestal niet meer dan maximaal tien tot vijftien, die weer zijn onderverdeeld in vergelijkbare aantallen subrubrieken, enzovoort. In eerste instantie zijn die vooral bedoeld om materiaal te ordenen en - zowel fysiek als virtueel - overzichtelijk neer te kunnen zetten, zodat je al browsend - langs boekenplanken of op het scherm - bij de juiste onderwerpen terecht komt. Maar ook in zoeksystemen kan daarvan vaak gebruik worden gemaakt om zoekacties – vooraf of achteraf – tot een bepaalde onderwerpsrubriek te beperken.

Ook hier geldt dat bij verschillende collecties verschillende soorten systematieken worden gebruikt. Zo maken openbare bibliotheken gebruik van SISO-codes (Schema voor Indeling van de Systematische catalogus in Openbare bibliotheken), gebruiken veel Nederlandse wetenschappelijke bibliotheken NBC-rubrieken (Nederlandse Basis Classificatie), zijn veel Amerikaanse collecties ingedeeld volgens de Library of Congress Classification (LCC) en wordt in veel andere landen gebruik gemaakt van de Dewey Decimale Classificatie (DDC). Daarnaast bestaan er echter nog honderden andere indelingsschema's. Om hier bij het zoeken gebruik van te kunnen maken moet je dus op een of andere manier weten welk systeem bij de betreffende catalogus of database wordt gebruikt.

Zoeken op andere specifieke velden

Op welke velden je verder nog kunt zoeken is voor een deel afhankelijk van de specifieke catalogus of database. Ook zijn niet alle verdere velden altijd even nuttig voor een gemiddelde zoekactie. Het belangrijkste daarvoor zijn verder nog:

- Zoeken op publicatiejaar. Zowel voor een “known-item search” als voor onderwerpsvragen biedt dit de mogelijkheid om je resultaat - in combinatie met onderwerpstermen - in te perken op de meest waarschijnlijke of meest gewenste tijdperiode. Bij veel zoeksystemen kunnen dit soort inperkingen in een apart zoekhokje aan de zoekvraag worden toegevoegd. Ook sommige webzoekmachines bieden zulke inperkingen in een advanced zoekscherm (niet bij Google!).
- Zoeken op taal. Biedt de mogelijkheid om zoekresultaten in te perken op de gewenste taal of talen. Bij veel zoeksystemen kunnen dit soort inperkingen in een apart zoekhokje, via uitklapmenu's of met aanvinkhokjes aan de zoekvraag worden toegevoegd. Veel webzoekmachines bieden zulke inperkingen in een advanced zoekscherm (zie bijvoorbeeld het in paragraaf 3.2.1 afgebeelde Google-scherm).
- Zoeken op documentsoort. Biedt de mogelijkheid om zoekresultaten in te perken op alleen boeken, video's, rapporten, scripties en dergelijke. Welke inperkingen mogelijk zijn hangt af van de aard van de betreffende collectie. Bij veel zoeksystemen kunnen dit soort inperkingen in een apart zoekhokje, via uitklapmenu's of met aanvinkhokjes aan de zoekvraag worden toegevoegd. Met webzoekmachines kun je je zoekvraag hoogstens inperken op het file-format van het betreffende document (webpagina, PDF, Word, Powerpoint, Excel-sheet e.d.; zie bijvoorbeeld het in paragraaf 3.2.1 afgebeelde Google-scherm).
- Zoeken op woorden uit de abstract. In veel databases kun je ook specifiek zoeken op woorden die in de samenvattingen van de oorspronkelijke publicaties voorkomen. Bij bibliografische databases is deze zoekmogelijkheid niet zo zinvol. Meer inhoudelijke tekst dan die samenvattingen zit daar namelijk nauwelijks in, zodat net zo goed (standaard) in alle velden gezocht kan worden. Bij een full-text bestand is het wel zinvol alleen in samenvattingen te zoeken, in plaats van in de volledige tekst. Je hebt dan namelijk een veel grotere kans dat de gevonden publicaties echt over het gewenste onderwerp gaan, omdat in de samenvatting alleen de kern wordt weergegeven van waar het document over gaat.

3.3 Zoekmethoden

Weten welke zoekmogelijkheden de meeste zoeksystemen bieden is nog maar een eerste stap op weg naar de uiteindelijk toe te passen zoekmethoden. Daarbij gaat het er om hoe je uiteindelijk concreet je zoekvraag gaat aanpakken. Hoe je aan zoekwoorden komt, hoe je die zoekwoorden combineert, welke verdere zoekmogelijkheden je daarbij toepast, hoe je je zoekvraag aanpast op basis van de resultaten die je in eerste instantie gekregen hebt, enzovoort. Op die aspecten gaan we in deze paragraaf nader in.

3.3.1 Vraaganalyse en zoekwoorden

Voordat je metterdaad gaat zoeken, dien je eerst de zoekvraag goed te analyseren. Soms heb je zelf al een preciezere vraagafbakening gemaakt, waarin je invalshoeken en nadere specificaties van je onderwerp hebt bepaald (zie paragraaf 3.1). Anders moet je de omschrijving van je zoekprobleem alsnog ontleden in de inhoudelijke componenten waaruit dat is opgebouwd.

Stel dat je iets wilt weten over “*de markt voor kranten op internet*”. Aan die vraag zijn tamelijk eenvoudig drie aspecten te onderscheiden. Het gaat over “**kranten**” en dan specifiek op “**internet**”, waarbij het alleen over de “**markt**” voor die kranten moet gaan. In plaats van over aspecten, wordt ook wel gesproken over componenten, facetten of concepten waaruit de zoekvraag bestaat. Dit heet daarom ook wel een “conceptuele analyse”. Na goed lezen van de omschrijving van het zoekprobleem is het, met wat gezond verstand en enig analytisch vermogen, voor de meeste vragen niet zo moeilijk om een dergelijke analyse te maken.

Behalve deze inhoudelijke analyse van het probleem zul je ook nog moeten onderzoeken of er meer formele randvoorwaarden zijn. Daarbij kun je denken aan:

- hoe recent moet de te vinden informatie zijn?
- moet het alleen in het Nederlands zijn of ook in andere talen?
- moet het om de situatie in Nederland gaan of ook in andere landen?
- wat voor soorten publicaties wil je hebben, officieel gepubliceerde artikelen, onderzoeksrapporten, nieuwtjes op internet?
- hoe veel informatie of hoe volledig overzicht is gewenst?

Ook dat zijn namelijk allemaal elementen die in de uiteindelijke uitwerking van de zoekvraag meegenomen moeten worden.

In dit stadium is het ook al nuttig om over geschikte zoekwoorden na te denken. Zijn de woorden waarmee je de concepten hebt omschreven al voldoende specifiek, niet te algemeen en niet te specialistisch. Je kunt daarbij ook al denken aan alternatieven of synoniemen. Als je ook informatie in andere talen zoekt, horen daarbij ook woorden die je in die talen zou gebruiken voor de betreffende concepten. Met webzoekmachines zul je namelijk alleen Engelstalige informatie kunnen vinden wanneer je ook Engelse zoektermen gebruikt. Voor databases waarin alle informatie (ook oorspronkelijk Nederlands- of Spaans-talige) standaard in het Engels is beschreven, zul je in elk geval ook Engelse termen moeten bedenken.

3.3.2 Bouwsteenmethode

Een veel gebruikte aanpak van zoekvragen komt direct voort uit het conceptueel analyseren van de vragen. De daarin onderscheiden concepten worden dan beschouwd als de bouwstenen waarmee de zoekvraag wordt opgebouwd. Vandaar de benaming “bouwsteenmethode”. Deze bouwsteenmethode berust sterk op de mogelijkheid Booleaans te kunnen zoeken.




Voor elke bouwsteen (dus voor elk concept) bedenk je in de voorbereiding op het echte zoekproces welke woorden daarvoor karakteristiek zijn. Het zijn de woorden waarvan te verwachten is, dat documenten waarin dat woord voorkomt, zeer waarschijnlijk aandacht besteden aan dat concept. Voor het concept “kranten” is dat in onderstaand schema al voor een deel uitgevoerd. Daarbij beperken we ons even tot alleen Nederlandse woorden. Uiteraard komt het woord *krant* in aanmerking, maar

ook documenten waarin het synoniem *dagblad* voorkomt, zullen over kranten gaan. Datzelfde geldt voor woorden als *ochtendblad* of *avondblad*, ook al zijn dat in feite specifiekere begrippen voor bepaalde soorten dagbladen. Voor een uitputtend onderzoek kun je zelfs nog een stap verder gaan, want ook documenten waarin *Volkskrant*, *Telegraaf* of *NRC-Handelsblad* ter sprake komen, gaan over kranten (ook al komen de woorden “krant” of “dagblad” daar toevallig niet in voor). Dus ook die namen – en namen van andere kranten – komen in principe als zoektermen in aanmerking.

Bij elk van de zo verzamelde termen moet ook nog worden bedacht of het meervoud, andere woordvormen, zoals werkwoorden, vervoegingen en verbuigingen daarvan en samenstellingen ook in de vraag moeten worden meegenomen. In veel gevallen kan dat eenvoudig worden afgedaan met een truncatie. Maar bij zoeksystemen waar dat technisch niet mogelijk is, moeten alle relevante varianten echt apart worden ingetikt. In onderstaand voorbeeld is voor de overzichtelijkheid wel gewoon met een truncatie-teken gewerkt. De zo verzamelde zoekwoorden voor een vraagconcept worden uiteindelijk met de Booleaanse OR-operator gecombineerd, om alles wat maar enigszins over kranten gaat bij elkaar te krijgen.

Op dezelfde manier wordt te werk gegaan met de andere vraagconcepten. Het nogal brede begrip “markt” is hieronder nog niet uitputtend uitgewerkt. Wel zijn woorden die iets te maken hebben met *concurrentie* daar als extra termen toegevoegd. Dit als voorbeeld dat je soms ook moet denken aan woorden die formeel geen synoniem zijn van je oorspronkelijke zoekwoord, maar die binnen het kader van je vraagstelling toch karakteristiek zijn voor dat aspect van je zoekvraag.

De resultaten van de bouwstenen worden vervolgens met AND gecombineerd. Dat heeft dus tot gevolg dat je alleen die documenten overhoudt, waarin uit elk van deze rijtjes woorden op zijn minst één woord voorkomt, zodat elk van je drie concepten in het document vertegenwoordigd is.

	kranten	internet	markt
OR 	krant* dagblad* ochtendblad* avondblad*	internet web www website*	markt concurrent*
	 AND		 AND

Mogelijke uitwerking van de zoekvraag “de markt voor kranten op internet”, beperkt tot alleen Nederlandstalige zoektermen

In systemen waar je een zoekvraag in één keer moet opgeven, luidt die dan:

(krant* OR dagblad* OR ochtendblad* OR avondblad*) AND (internet OR web OR www OR website*) AND (markt OR concurrent*)

In een systeem dat stapsgewijze met resultaatsets werkt, kan dit worden:

#1 krant* OR dagblad* OR ochtendblad* OR avondblad*
#2 internet OR web OR www OR website*
#3 markt OR concurrent*
#4 #1 AND #2 AND #3

3.3.3 Best match zoeken

De bouwsteenmethode is een heel systematische en overzichtelijke methode, waarmee je vragen heel precies kunt specificeren. Toch is hij niet voor elke situatie geschikt. Niet bij elke vraag is het zo eenvoudig hem op zo'n conceptueel analytische manier te ontleden. En ook zal niet elke “zoeker” zal van nature zo'n conceptueel-analytische denkwijze hanteren. In dergelijke gevallen is toepassing van de in paragraaf 3.2.1 beschreven techniek van “best match” zoeken een goed alternatief. Die levert dan bijna altijd wel enige relevante resultaten op.

Over het algemeen werkt deze methode beter naarmate je meer en specifiekere zoekwoorden intikt. Wel zal het dan soms nodig zijn om, in plaats van op *all of the words*, te zoeken op *any (of at least one) of the words*. Een precieze ordening van je zoekwoorden over drie of vier basisconcepten is daarbij niet nodig. Ze kunnen gewoon achter elkaar worden ingetikt.

Anderzijds is het moeilijk met deze methode in één keer een tamelijk volledig overzicht van relevante informatie te krijgen. Daarvoor zul je meestal verschillende reeksen woorden moeten uitproberen, met daarin telkens andere synoniemen van je oorspronkelijke zoekwoorden. Bij “pure” toepassing van deze methode kun je immers niet aangeven dat bepaalde woorden met OR gecombineerd moeten worden, omdat het synoniemen zijn. In de praktijk blijkt dit, vooral bij zoekmachines, gelukkig wel vaak mogelijk, zoals onderstaand “advanced” zoekscherm van Google illustreert.

Voorbeeld van “best match” zoekactie waarbij toch rekening wordt gehouden met het combineren van de synoniemen “krant, kranten, dagblad, dagbladen”.

In feite is deze zoekactie equivalent met de Booleaanse combinatie

web AND concurrentie AND (krant OR kranten OR dagblad OR dagbladen)

waarbij dan nog wel relevantieordering wordt toegepast, onder meer op basis van waar en hoe vaak deze zoekwoorden in de gevonden webpagina's voorkomen.

Nadeel van de “best match” methode is ook dat hij (zeker voor gebruikers die wel heel conceptueel denken) wat ondoorzichtiger is en dat het wat moeilijker is om hem heel gestructureerd en planmatig toe te passen. Anderzijds is een groot voordeel van deze methode –zeker bij zoeken in grote informatiecollecties zoals het “hele” web – dat het systeem in feite een voorselectie voor de gebruiker maakt. Bij 17235 zoekresultaten, zullen gebruikers namelijk toch nooit meer dan hooguit de eerste twintig of dertig daarvan goed kunnen en willen bekijken, zodat het van belang is dat de belangrijkste en meest relevante bovenaan de lijst staan. Daarvoor moeten dan dus heel goede technieken voor relevance ranking worden toegepast.

3.3.4 Uitbreiden van zoekresultaten

Een wezenlijk aspect van digitale zoekacties is dat zoekvragen zo makkelijk kunnen worden aangepast en gevarieerd. Meestal gebeuren dat soort aanpassingen op basis van een beoordeling van al verkregen resultaten. Soms zal die beoordeling heel oppervlakkig gebeuren; je ziet al in één oogopslag dat je helemaal fout zit en het anders moet aanpakken. In andere gevallen zul je de gevonden resultaten veel zorgvuldiger bekijken, bijvoorbeeld om te zien of die je op ideeën brengen om andere, betere zoekwoorden te gebruiken.

Wanneer je zoekt in collecties die al tamelijk gespecialiseerd zijn en betrekkelijk klein, zeker in vergelijking met het hele internet, zul je vaak de behoefte hebben je zoekresultaat uit te breiden. Ondanks de goede voorbereiding van je zoekvraag, heb je het idee dat er beslist meer en misschien ook nog wel betere informatie te vinden moet zijn. En zelfs bij gebruik van een webzoekmachine bestaat die behoefte nog vaak. We zullen hier een aantal verschillende manieren bekijken waarop je dat kunt doen.

De eenvoudigste manier om het zoekresultaat bij de bouwsteenmethode te vergroten, is om een concept helemaal weg te laten uit je zoekvraag. Dat is natuurlijk gevaarlijk, want je had niet voor niets vooraf een conceptuele analyse gemaakt. Een belangrijk aspect daaruit weglaten, leidt er vaak toe dat veel te veel van wat je dan vindt helemaal niet relevant is. Toch komt het nogal eens voor dat een bepaald concept bij nader inzien overbodig blijkt. Het kan zijn dat het eigenlijk al impliciet besloten ligt in de rest van je zoekvraag, of dat de gespecialiseerde collectie waarin je zoekt, toch alleen maar documenten op dat betreffende onderwerpsterrein bevat. Als je voor de krantenzoekvraag uit paragraaf 3.3.2 een zoekactie uitvoert in een database met alleen marketinginformatie, zul je de bouwsteen voor het concept "markt" waarschijnlijk straffeloos kunnen weglaten. De relevantie van zo gevonden documenten zal daardoor nauwelijks nadelig beïnvloedt worden, omdat bijna alles wat in die database zit, al met markt-aspecten te maken heeft, ook de documenten waarin jouw oorspronkelijke zoekwoorden *markt* of *concurrentie* niet expliciet voorkomen. Bij een best match zoekactie zal het weglaten van een zoekwoord over het algemeen op dezelfde manier tot een uitbreiding van het zoekresultaat kunnen leiden.

Bij de voorgaande methode hoefde je nog niet zo erg naar je zoekresultaat te kijken, al kan dat soms wel helpen om te bedenken welk concept je zou kunnen weglaten. Bij de volgende manieren is dat zoekresultaat wel steeds het uitgangspunt. Je zou ze daarom allemaal kunnen karakteriseren met de omschrijving "sneeuwbal-methode". Je hebt al wat gevonden en op basis daarvan vind je nog meer.

Bij het eerste type sneeuwbal probeer je aan al gevonden resultaten ideeën te ontlenu voor extra zoektermen. Als je bij de voorbereiding van je zoekvraag bijvoorbeeld nog niet had bedacht dat *ochtendbladen* en *avondbladen* ook mogelijke zoektermen zijn voor het concept *kranten*, dan kun je alsnog op dat idee gebracht worden doordat in de documenten die je met de zoektermen *krant* of *dagblad* vindt, toevallig ook deze woorden blijken voor te komen. Zo worden de bouwstenen dus achteraf aangevuld met woorden die je zelf nog niet bedacht had. Zo kun je echte synoniemen tegenkomen, of specifiekere deelbegrippen (zoals die twee specifieke types dagbladen), of de namen van specifieke voorbeelden (zoals namen van bepaalde kranten), of gespecialiseerd jargon wat je zelf eigenlijk niet kende. In sommige gevallen zal het ook een woord zijn dat bij nader inzien veel beter is dan een eerder gebruikt zoekwoord en ter vervanging daarvan kan dienen. Zorg hierbij wel dat je in je oorspronkelijke zoek-

schema - het tabelletje met zoekwoorden voor je bouwstenen - vastlegt hoe je dat schema successievelijk uitbreidt. Dat dwingt je ook voor elke nieuwe zoekterm te bedenken bij welk concept of bouwsteen het thuishoort, zodat je het op de goede manier in de OR- en AND-combinaties verwerkt. Op een vaak wat minder gestructureerde manier, kun je deze techniek in principe ook toepassen bij een best match zoekactie.

Een tweede type sneeuwbal is vrijwel alleen toe te passen wanneer een zoekactie boeken of artikelen heeft opgeleverd, waar in de gevonden resultaten zichtbaar is wie de auteurs daarvan zijn. Als je daarbij een bijzonder belangwekkende of relevante publicatie hebt gevonden, kun je kijken of dezelfde auteur nog meer heeft geschreven over dat onderwerp of over verwante thema's, door simpelweg alleen op die auteursnaam verder te zoeken. Op dezelfde manier kun je ook instituten, organisaties of bedrijven op het spoor komen die voor jouw onderwerp belangrijk zijn en op hun website wellicht nog meer en andere informatie daarover hebben staan.

Een derde type sneeuwbal is weer heel algemeen toepasbaar. Veel artikelen bevatten literatuurreferenties die verwijzen naar andere, al eerder verschenen publicaties, waarop wordt voortgeborduurd, die worden tegengesproken of waarvan anderszins gebruik wordt gemaakt. Die zullen dus altijd een relatie met het oorspronkelijke onderwerp hebben, zodat het nuttig kan zijn om die ook op te sporen. Dat kan meestal op basis van in de verwijzing vermelde gegevens over auteur, titel en vindplaats. Dat worden dan dus "known-item searches". Bij gevonden webpagina's gaat dit meestal veel makkelijker. Daarin komen namelijk hyperlinks voor, waarvan de functie vaak vergelijkbaar is met die van literatuurreferenties in artikelen. Dergelijke links hoef je alleen aan te klikken om op die andere gerelateerde webpagina's terecht te komen.

Een vierde type sneeuwbal tenslotte is het omgekeerde hiervan. Als je een relevante publicatie hebt, dan zullen ook latere publicaties die daaraan refereren, waarschijnlijk relevant zijn. Alleen is aan die uitgangspublicatie zelf niet te zien welke dat zijn, want de auteur daarvan kon uiteraard niet in de toekomst kijken, wat later nog met zijn ideeën gedaan zou worden. Daarvoor heb je dus speciale hulpmiddelen nodig. Die bestaan vrijwel alleen voor wetenschappelijke en vakpublicaties. De belangrijkste daarvan zijn de al ruim 50 jaar bestaande *Citation Indexes* op het gebied van de alfa-, beta- en gamma-wetenschappen, de heel recente Scopus-database van Reed-Elsevier en het gratis wetenschappelijke zoekstelsel Google Scholar. Wanneer je in Google Scholar een zoekactie hebt gedaan, staat bij elke gevonden publicatie vermeld hoe vaak die door andere in het systeem aanwezige publicaties wordt "geciteerd" ("cited by" betreft hier literatuurverwijzingen; het betekent niet dat er letterlijk geciteerd is). Die citatie-vermelding is een link die je alleen hoeft aan te klikken om een lijstje te krijgen van (recentere) publicaties die naar het uitgangsverhaal verwijzen. Ook bij gewone webpagina's zijn deze omgekeerde relaties te volgen met behulp van een aantal van de bekende zoekmachines. Met de zoekvraag

link:<http://www.website.org/directory/interessant.html>

vind je webpagina's die een link bevatten naar het daar vermelde URL. Als dat het URL van een tamelijk gespecialiseerde pagina is, zal daar ook vaak een inhoudelijke relatie tussen zijn. Bij onder meer Yahoo, AltaVista en Google is deze zoekmethode mogelijk.

The screenshot shows a Google Scholar search for 'information seeking behavior'. The search results page displays the first 10 results out of approximately 228,000. The top result is 'Dimensions of Consumer Information Seeking Behavior' by G.C. Kiel, published in the Journal of Marketing Research in 1981. Below this, several other articles are listed, including 'Information-Seeking Behavior' by J. Krikelas, 'Information seeking behavior group of 7' by C.M. Brown, 'Cancer patients' information needs and information seeking behavior' by G.M. Leydon et al., and 'A Model of Tourist Information Search Behavior' by D. Fodness and B. Murray. A second search box is overlaid on the page, showing a search for 'Kiel: Dimensions of Consumer Information Seeking Behavior', which returns 74 citing articles, including 'Premarket Forecasting of Really-New Products' and 'A Dyadic Study of Interpersonal Information Search'.

Citatie-relaties bij Google Scholar: link naar de 74 artikelen die een verwijzing bevatten naar het artikel van G.C. Kiel uit 1981

The screenshot shows an AltaVista search for the URL 'http://www.scils.rutgers.edu/~sympark/colis.html'. The search results page displays 9 results. The first result is 'Information Seeking Behavior' from librarysupportstaff.com. The second result is 'WebWord " Blog Archive " The Psychology of Search: Chapter Three' from webword.com. The third result is 'LBR 200: Information Literacy, Fall 2004' from library.umaine.edu. The search interface includes a search bar with the URL, a 'FIND' button, and options for search scope (Worldwide, USA) and language (All languages, English, Spanish).

Met "link:"-search vind je pagina's die een link bevatten naar het daarachter ingevulde URL

3.3.5 Inperken van zoekresultaten

Bij het doen van zoekacties bestaat ook heel frequent de behoefte om zoekresultaten in te perken. Die behoefte kan domweg voortkomen uit het feit dat je zo ontzettend veel hebt gevonden. Als een zoekstelsel die 17235 resultaten dan niet op relevantie ordent, zul je zelf een zinnige inperking moeten maken. Het selecteren van de vijftig meest recente is natuurlijk een heel pragmatische aanpak, maar er zijn ook meer inhoudelijke manieren. In andere situaties zal die behoefte vooral worden ingegeven door een kwaliteitsoordeel: veel te veel van wat je hebt gevonden, blijkt onvoldoende - of zelfs helemaal niet - relevant voor je onderzoeksvraag. Bij het inperken zullen dan zeker inhoudelijke argumenten moeten meespelen. Inperkingen zullen vaak al onmiddellijk op de allereerste zoekresultaten worden toegepast, maar ook nadat een zoekactie is uitgebreid zullen daar vaak weer inperkingen op volgen.

Inhoudelijke inperkingen kunnen uiteraard worden uitgevoerd door iets te veranderen in de oorspronkelijk gebruikte zoekwoorden. Als je aan de zoekresultaten ziet dat je vraag eigenlijk nog onvoldoende is gespecificeerd of afgebakend, kun je - in een bouwsteen-aanpak - proberen een extra concept aan de zoekvraag toe te voegen. Een AND-relatie met dat concept zal het zoekresultaat meestal flink inperken naar aantal en de relevantie ervan sterk verbeteren. Bij "best match" zoeken heeft toevoegen van één extra zoekwoord meestal al het gewenste effect. Het bedenken van een inaanmerking komend extra concept of zoekwoord kan voortkomen uit een nadere analyse van de zoekvraag zelf, maar ook hiervoor kan een zorgvuldige scan van de gevonden resultaten je op ideeën brengen. Bij de eerdere krantenvraag zou je je op grond daarvan bijvoorbeeld kunnen realiseren dat het er eigenlijk vooral om gaat hoe de krant ook op internet interessant kan zijn voor potentiële *adverteerders*.

The screenshot shows the Ask.com search interface. At the top, there's a search bar with 'ajax' entered and a 'Search' button. Below the search bar, there's a red banner with 'Web Search' and 'Showing results 1-10 of 5,214,000'. The search results are organized into several sections:

- Search Tools:** Look up [Ajax, Ontario](#) | [Weather](#) | [Map](#) | [Time](#) | [Business Listings](#)
- Encyclopedia: Ajax Amsterdam** (Source: Wikipedia)

Amsterdamsche Football Club **Ajax** (Euronext: **AJAX**) also referred to as **Ajax Amsterdam**, **AFC Ajax**, or simply **Ajax** (pronounced *Ah-yahx*), is a football club from Amsterdam, Netherlands. The club is historically one of the top-10 football clubs in the world, and one of the three ... [More »](#)

Other Matches:
- Latest News: Ajax**

A Ghost image !!! [WebReference.com](#) 1 hour ago
- Sponsored Results:**
 - [Share your mashups](#) - Publish apps for use with blogs **Ajax**, Flash, Javascript, etc. [www.widgetbox.com](#)
 - [Beyond AJAX](#) - Pure Java RIA with Canoo UltraLightClient [www.canoo.com/ulc](#)
 - [Build Thin Client GUIs](#) - Custom Interactive JSF Web Apps View on demand technical webinars [www.ILOG.com](#)
- Narrow Your Search**
 - [Ajax Greek](#)
 - [Ajax Football Club](#)
 - [Ajax Cleaner](#)
 - [Ajax And Mythology](#)
 - [Suicide of Ajax](#)
 - [Ajax Football Club Holland](#)
 - [Greek Hero Ajax](#)
 - [Ajax Amsterdam Fc](#)
 - [Ajax Soap](#)
 - [Ajax Dish Soap](#)
 - [Greek God Ajax](#)
 - [AFC Ajax](#)
 - [Ajax Ontario](#)
 - [Ajax Football Team](#)

Zoekmachine Ask presenteert (rechts) suggesties om een zoekresultaat in te perken

Er zijn ook zoeksystemen, bijvoorbeeld de webzoekmachine Ask, die zelf al met suggesties komen, waarmee je het zoekresultaat zou kunnen inperken. Dat gebeurt vaak op basis van een statistische analyse van in je zoekresultaat voorkomende woorden. Ook eerder door andere gebruikers gestelde preciezere zoekvragen worden wel als bron daarvoor gebruikt.

Behalve deze manier, waarbij je preciezer specificeert waar te vinden informatie WEL over moet gaan, kun je natuurlijk ook aangeven waar het beslist NIET over mag gaan. Dat kan door woorden met de NOT-operator uit te sluiten. Dat uitsluiten van termen moet wel met voorzichtigheid gebeuren, omdat je ongewild juist ook relevante informatie kunt kwijtraken. Als je bij de eerdere krantenzoekvraag niet geïnteresseerd was in de gratis bladen Metro en Spits, is het toch niet verstandig die met NOT uit te sluiten. Je raakt dan namelijk ook alle publicaties kwijt waarin de invloed van die gratis bladen op de markt voor de gewone kranten aan de orde komt.

Bij de bouwsteenmethode kun je het zoekresultaat ook inperken en verbeteren zonder een extra concept toe te voegen of termen met NOT uit te sluiten. Als je goed naar je zoekresultaat kijkt, blijkt soms dat het vooral een bepaalde term uit één van de bouwstenen is, die aanleiding geeft tot veel ongewenste resultaten. Dat kan zijn omdat het een veel te algemeen woord is of omdat het te vaak in een niet bedoelde betekenis wordt gebruikt. In de krantenzoekactie zou bijvoorbeeld kunnen blijken dat in de "markt"-bouwsteen het woord "markt" zelf erg vaak wordt gebruikt zonder dat echt de markt voor het uitgeven van kranten wordt bedoeld. Dan is het raadzaam dat woord alsnog uit de OR-relatie van die bouwsteen weg te laten.

Andere manieren van inperken vallen meer onder de zoektechniek. Door je zoekactie te beperken tot alleen woorden uit titels van publicaties of tot trefwoorden waarmee publicaties gekarakteriseerd zijn, wordt je zoekresultaat sterk ingeperkt (zie 3.2.4). Daarbij zal de relevantie zeker verbeteren, omdat publicaties waar je zoekwoorden in de titel voorkomen of waaraan die woorden bewust als trefwoord zijn toegevoegd, een veel grotere kans hebben dat onderwerp echt als centraal thema te hebben.

Als een deel van je zoekwoorden in de praktijk meestal als vaste uitdrukking zal voorkomen, kun je ook op "exacte zinnen" zoeken. Er blijven dan alleen resultaten over waar de gevraagde woorden exact in de ingetikte volgorde voorkomen, zonder andere woorden ertussen. Dat is dus een drastische extra eis. Bij webzoekmachines doe je dat vrijwel altijd door het gewenste zinnetje tussen aanhalingstekens te zetten. Bij toepassing van de bouwsteenmethode is dit soms nogal lastig, omdat er zo veel mogelijke combinaties kunnen zijn van de in elke bouwsteen verzamelde woorden. In die gevallen is het een voordeel wanneer in een zoekstelsel de in paragraaf 3.2.2 besproken afstandsoperatoren gebruikt kunnen worden. In zo'n stelsel zou de krantenzoekvraag bijvoorbeeld kunnen luiden

(krant* OR dagblad* OR ochtendblad* OR avondblad*) NEAR (internet OR web OR www OR website*) NEAR (markt OR concurrent*)

om aan te geven dat woorden uit de drie bouwstenen niet alleen samen moeten voorkomen, maar ook dicht bij elkaar in de buurt moeten staan. Dat is veel preciezer dan met AND, maar al weer een stuk minder drastisch dan de eis van exacte zinnetjes. Verder moet een zoekvraag vaak nog worden ingeperkt op mogelijke formele kenmerken die in paragraaf 3.3.1 als mogelijke randvoorwaarden werden genoemd: recentheid, taal, soort publicatie en dergelijke. Ook daarmee beperk je de grootte van het resultaat en focus je beter op datgene waarnaar je op zoek bent.

4 Zoekplan

In de voorgaande paragrafen zijn de belangrijkste basisingrediënten aangedragen die nodig zijn om succesvol informatie te zoeken. Kennis van soorten informatiebronnen, van informatiecollecties en zoekhulpmiddelen, van zoektechnieken en zoekmethodes. In de juiste combinatie moeten die uiteindelijk tot bevredigende resultaten leiden. Gezien de vele keuzes die je in het hele zoektraject moet maken, moet daar wel een zoekplan aan ten grondslag liggen. Zeker voor een heel diepgaand informatieonderzoek, maar toch ook wel – in eenvoudiger vorm – voor wat oppervlakkiger of oriënterend onderzoek. Een dergelijk zoekplan is in feite een soort “plan van aanpak”, zoals je dat voor elke activiteit, elk onderzoek of elk project moet maken. Een zoekplan is echter wel specifiek toegespitst op de stappen die bij een informatieonderzoek een rol spelen. In feite is het een stappenplan. De opeenvolging van al die stappen mag je dan de uiteindelijke zoekstrategie noemen. Zoals bij elk “plan van aanpak” geeft het zoekplan structuur aan het hele zoektraject, biedt het duidelijkheid en kadert het in.

4.1 Stap 1: Afbakening en aard van de vraag

In paragraaf 3.1 werd al ingegaan op het belang van een goede afbakening van je zoekvraag. Daar werd ook een voorbeeld gegeven. Dat illustreerde dat je meestal vanuit een breed kader beginnend, in een soort trechtersvorm, de vraag steeds specifieker zult omschrijven. Zo kom je als het ware tot een steeds preciezer “filter”, dat in een ideale situatie alleen nog relevante informatie zou moeten doorlaten.

Naast deze inhoudelijke afbakening dien je vast te leggen wat de aard en de bedoeling van de vraag zijn. Of het oriënterend of diepgaand is, hoe volledig het zoekresultaat moet zijn, van welk niveau de te vinden informatie moet zijn, wetenschappelijk of juist meer populair, of heel recente of juist ook oudere informatie nodig is, enzovoort. Wat het doel van de informatie is, of het nodig is voor een strategische beslissing (zoals financiering of overname van een bedrijf), voor een concurrentieanalyse of voor een reclamecampagne, als achtergrond voor een tv-interview of voor een documentaire, voor interpretatie van een eigen onderzoek of voor het schrijven van een degelijk achtergrondartikel voor een kwaliteitskrant.

Voorts is het nuttig te inventariseren wat je al weet:

- Ken je al publicaties over het onderwerp?
- Ken je auteurs die hierover gepubliceerd hebben?
- Ken je mensen die er verstand van hebben?

Zolang de vraagstelling niet duidelijk gespecificeerd is en die achtergrondvragen niet zijn beantwoord, kunnen de volgende stappen in het zoekproces nog niet worden uitgevoerd en zelfs nog niet in detail worden gepland. Deze stap moet dus al zijn uitgevoerd voor de rest van het zoekplan in meer detail kan worden uitgewerkt.

4.2 Stap 2: Te gebruiken bronnen en systemen

De inhoudelijke afbakening en de aard en bedoeling van de zoekvraag vormen de basis voor de keuze van te gebruiken bronnen en zoeksystemen. De aard van de informatievraag bepaalt of we op zoek gaan naar “literatuur over”, naar harde

gegevens, naar statistieken, naar productinformatie e.d. Dat bepaalt ook of we krantenberichten als bron willen gebruiken of dat we in bronnen met wetenschappelijke literatuur moeten zoeken, of we boeken met uitgekristalliseerde kennis en opvattingen willen vinden of artikelen met de laatste misschien nog speculatieve onderzoeksresultaten. Op grond daarvan bepalen we in welke informatiecollecties we verwachten de grootste kans op succes te hebben om de gezochte soort informatie te vinden. En zo kunnen we uiteindelijk een keuze maken voor de zoeksystemen die we hiervoor willen gebruiken.

In het zoekplan wordt al zo goed mogelijk vastgelegd hoe je dit denkt aan te pakken. Uiteraard zul je in de uitvoering van het zoekplan flexibel moeten zijn. Wanneer je daarbij merkt dat bepaalde bronnen en zoeksystemen niet het resultaat opleveren dat je gehoopt had, terwijl een kijkje in een systeem waarvan je geen hoge verwachtingen had toch veelbelovend blijkt, dan zul je uiteraard je plan alsnog moeten aanpassen. Overigens zouden in een heel degelijk zoekplan dergelijke "noodscenario's" zelfs al voorzien moeten zijn.

4.3 Stap 3: Te gebruiken zoekwoorden

De afgebakende vraagstelling uit stap 1 moet je vertalen in concepten, die je bijvoorbeeld in een bouwsteenmethode zou kunnen gebruiken (zoals beschreven in paragraaf 3.3.1). Bedenk bij die concepten al mogelijke zoekwoorden. Denk daarvoor aan synoniemen, aan specifiekere deelbegrippen en bijvoorbeeld ook aan namen van producten, technieken, personen, bedrijven e.d. die kenmerkend kunnen zijn voor onderdelen van je zoekvraag. Vertaal de woorden zo nodig ook in andere talen wanneer je ook informatie in die talen wilt vinden. Ook in dit deel van je zoekplan zullen tijdens de uitvoering van je zoekproces nog aanvullingen en wijzigingen aangebracht kunnen worden.

4.4 Stap 4: Te gebruiken zoekmethode(s)

Bepaal, mede op basis van je analyse in stap 3, of de vraag zich wel leent voor een bouwsteenmethode (zoals besproken in paragraaf 3.3.2) of dat in dit geval beter een best-match zoekactie (paragraaf 3.3.3) kan worden uitgevoerd. In het eerste geval moeten de in de vorige stap bedachte zoekwoorden zorgvuldig bij de juiste concepten of bouwstenen worden ingedeeld.

Probeer ook hier al te bedenken welke acties je zou kunnen ondernemen om je zoekresultaat in te perken, in het geval je veel te veel vindt en de relevantie van het gevonden materiaal erg zou tegenvallen (zie paragraaf 3.3.5). Bedenk anderzijds ook al welke methoden in aanmerking komen in het geval je juist meer wilt vinden. Voorzie je al dat daartoe een bepaald zoekconcept uit de vraag kan worden weggelaten. Of komen wellicht bepaalde sneeuwbal-technieken in aanmerking, omdat je al een goede publicatie hebt die je daarvoor als uitgangspunt kunt nemen (zie paragraaf 3.3.4).

Ook hier geldt dat een goede planning en voorbereiding weliswaar het halve werk is, maar dat de uitvoeringspraktijk weerbarstig kan zijn waardoor je je oorspronkelijke planning zult moeten aanpassen. Zoeken is nu eenmaal een in hoge mate interactief proces, zodat je die flexibiliteit moet hebben.

4.5 Stap 5: Beoordeling van de resultaten en selectie van de informatie

In de voorgaande stappen is al aangegeven dat je ook aanpassingen aan je zoekplan moet voorzien. Tijdens de werkelijke uitvoering van je zoekplan zul je dus al voortdurend je resultaten beoordelen om te zien of je zoekacties inderdaad opleveren wat je gehoopt had, of dat inderdaad aanpassingen nodig zijn. Daarnaast zul je in het zoekplan vaak ook al proberen aan te geven hoe je te werk zult gaan om uit de gevonden artikelen, boeken, websites, rapporten, projectbeschrijvingen en dergelijke, uiteindelijk die publicaties of die gegevens te selecteren die je nodig hebt. De gegevens die je in je rapportage wilt verwerken, de publicaties waarnaar je daarin wilt verwijzen, de projecten die je wilt noemen. In de praktijk zal de kwaliteit van de gevonden informatie daarbij ook een grote rol spelen. Aan methoden en kenmerken voor het beoordelen van de kwaliteit en de betrouwbaarheid van gevonden informatie, hebben we in deze tekst geen aandacht besteed. Dat is een heel afzonderlijk onderwerp dat in andere publicaties uitgebreid aan de orde komt.

4.6 Stap 6: Evaluatie van het zoektraject

Bedenk ook al hoe je na afloop het hele zoektraject zult evalueren. Wie betrek je erbij om het eindresultaat te beoordelen? Hoe beoordeel je of alle keuzes die je gedaan hebt in je zoekplan en in de concrete uitvoering van je onderzoek gerechtvaardigd waren en het verwachte effect hadden.

4.7 Vastlegging van het zoektraject

In de meeste gevallen is het van belang om de uitvoering van het hele zoektraject te documenteren. Leg bij de daadwerkelijke uitvoering van alle stappen uit je zoekplan, bijvoorbeeld in een soort logboek, vast dat en hoe ze zijn uitgevoerd. Daarin komen dan zaken te staan zoals:

- in welke bronnen of systemen gezocht is,
- wanneer dat werd gedaan,
- welke zoekvragen en combinaties van zoekwoorden zijn gebruikt,
- wat die opleverden (ook als het niets opleverde),
- hoeveel daarvan nuttig of relevant was voor het onderzoek (ook als niets relevant bleek),
- hoe de zoekvraag werd aangepast,
- wat dat voor effect had,
- en dergelijke.

Zo'n verslaglegging biedt je de mogelijkheid om achteraf nog eens te kijken of je het wel goed had aangepakt en of je geen zinvolle mogelijkheden over het hoofd hebt gezien. Ook kun je op grond hiervan je zoekactie later nog eens overdoen – in de hoop dat intussen nieuwe informatie beschikbaar is gekomen – waarbij je de valkuilen die je eerder was tegengekomen kunt proberen te ontlopen.

[Voor deze paragraaf is onder meer gebruik gemaakt van hoofdstuk 4.1 uit het boek Informatievaardigheden (Informatievaardigheden / Albert Boekhorst; Inge Kwast, Diane Wevers. – 3^e dr. – Utrecht: Lemma, 2004). De hier gebruikte indeling in stappen wijkt echter enigszins af van het schema waarop Boekhorst uitkomt.]