

HET ZOEKEN VAN WETENSCHAPPELIJKE INFORMATIE OP INTERNET

Drs Marie José Vlaanderen, MSc

Universiteitsbibliotheek
Erasmus Universiteit Rotterdam
augustus 2001

Verantwoording	3
1 Inleiding	4
2. De gebruiker	6
3 Zoekmachines	8
3.1 De gewone zoekmachines	8
3.2 Portals	9
4. De betere zoekmachines	9
5. Subject Gateways	10
5.1 Inleiding	10
5.2 DESIRE	11
5.3 Evaluatie	12
6 Intelligent agents	14
6.1 Inleiding	14
6.2 InfoSpiders	15
6.3 Kenjin	16
6.4 Shopping agents	16
7. Resultaten	17
7.1 Conclusie	17
7.2 Uitkomsten naar aanleiding van de vraagstelling	18
7.3 Voorstellen voor een UB-helppagina voor Internet	19
8. Referenties	20
APPENDIX Uitslag enquête	22

Verantwoording

Dit onderzoek is uitgevoerd in opdracht van de Universiteitsbibliotheek (UB) van de Erasmus Universiteit Rotterdam (EUR) in de periode oktober 1999 – april 2000.

De bedoeling is een overzicht te geven van de stand van zaken (*state of the art*) op het gebied van zoekfaciliteiten die gebruikt worden of kunnen worden bij het zoeken naar wetenschappelijke informatie op het Internet, in het bijzonder voor studenten en medewerkers van de EUR.

Er is gebruik gemaakt van kennis uit (tijdschrift-)literatuur uit de jaren 1998-2000, Internetrapporten zoals die te vinden zijn bij www.scout.com, www.brint.com, www.searchenginewatch.com, lezingen op het CIKM'99¹-congres (Kansas City, Mo, 1999), een cursus zoeken op Internet, en de resultaten van een enquête die gehouden is onder medewerkers van de faculteiten op Woudestein in maart 2000.

¹ Conference on Information- and Knowledge Management

1 Inleiding

Internet heeft in 2000 de miljardste site binnengehaald. Er is geen sprake van enige structuur, conformiteit, of iets dat anderszins ons houvast kan geven als we informatie van het World Wide Web (verder *het Net* of *Internet*² genoemd) willen halen. Chaos is de beste omschrijving.

Toch is het redelijk mogelijk een weg te vinden in die chaos als we maar de juiste hulpmiddelen gebruiken. In deze studie zijn die hulpmiddelen onderzocht. Er is niet getracht uitgebreide overzichten van zoekmachines, bronnenlijsten, e.d. te geven; deze zijn elders in overvloed aanwezig.

Ondanks de euforie dat *alles op het Net te vinden is* (wat overigens absoluut niet waar is) moeten we de beperkingen van het zoeken en vinden van informatie op het Internet vooral niet vergeten.

We moeten ook het Internet in de juiste verhouding zien met de reeds bestaande media. Vergelijk de zes miljoen wetenschappelijke sites (zie hieronder bij punt 2) die Internet te bieden heeft met de tien miljoen referenties die alleen al de databank Medline aanbiedt, of met de 12,5 miljoen referenties van het Web of Science. Bovendien zijn deze databanken goed gestructureerd en kan zo gezocht worden dat er vrijwel niets gemist hoeft te worden.

In zijn keynote lezing *Searching the WEB: can you find what you want?* op het CIKM'99 gaf C.Lee Giles een opsomming van de beperkingen van Internet, juist ook voor de wetenschapper.

1. Er is nog relatief weinig via het Net beschikbaar, in vergelijking met wat er aan boeken, tijdschriften, archieven, e.d., beschikbaar is.
2. 83% van wat op het Net beschikbaar is, is commerciële informatie. Van de overige 17% is maar 6% wetenschappelijk, de andere 11% is overheid, porno, religie, e.d. (zie ook Lawrence en Gilles 1999).
3. 65% van de sites zijn Engelstalig (dus is 35% voor de meeste mensen onleesbaar).
4. Search engines³ hebben een beperkte capaciteit. Het automatisch indexeren van de 800 miljoen (en inmiddels meer) webpagina's kost een maand tijd. Doordat rijp en groen door elkaar geïndexeerd wordt, is het resultaat vaak onbruikbaar.

Giles verwacht dat pas over tien jaar methodes ontwikkeld zullen zijn om goede zoekacties uit te kunnen voeren. Er zijn al wel speciale *search companies* die handmatig in opdracht zoeken.

² Er is wel verschil (bijv. HTML versus FTP) maar dat is voor de gebruiker niet belangrijk.

³ Zoekmachines, alhoewel juister is van indexermachines te spreken.

Eveneens tijdens het CIKM'99 gaf R. Rastogi (Lucent Bell Co) een lezing over *Data mining and the Web*. Data mining⁴ in grote databanken is een goed ontwikkelde techniek met goede en betrouwbare resultaten. Op het Net is datamining heel wat lastiger. Om te beginnen is de aangeboden structuur volstrekt willekeurig, verder zijn er de hyperlinks, de voortdurende veranderingen en willekeurig gedrag van gebruikers. (Niet genoemd is dat data mining in de eigenlijke betekenis alleen mogelijk is in een relationele databank.) Het is een uitdaging om in de komende jaren algoritmen voor data mining te ontwikkelen die op het Net losgelaten kunnen worden. Zover is het echter nog niet.

Wat betreft information retrieval⁵ noemt Rastogi dezelfde beperkingen die we reeds weten: *99% van de informatie is niet van belang voor 99% van de webgebruikers*.

Het verbeteren van deze beperkingen moet gezocht worden in betere web directories⁶. Rastogi noemt Clever, de automatische web directory ontwikkelaar van IBM (zie onder de betere zoekmachines) en Inktomi Directory Engine die concept inductie⁷ toepast. Clever is gebaseerd op het ontdekken van *web communities*. Het principe is om hyperlinks te ontdekken die frequent voorkomen (*hubs*) en naar autoriteiten (zoals sites van universiteiten) verwijzen. Door een bepaald algoritme toe te passen zijn per onderwerp betrouwbare *hubs* en autoriteiten te ontdekken. Zo wordt het kaf van het koren gescheiden. Dit principe lijkt veelbelovend, voorlopig de enige methode die al in een redelijk ver stadium van ontwikkeling is.

Dit verslag heeft twee invalshoeken. Ten eerste de menselijke factor: wat verwacht de gebruiker van Internet; wat voor informatie wordt gezocht; wat zijn typische knelpunten. De andere invalshoek is de techniek die voor oplossingen zorgt. In de hoofdstukken over techniek worden zoekmachines, subject gateways⁸ en intelligent agents⁹ behandeld. Aan dit laatste onderwerp is wat uitgebreider aandacht besteed omdat de verwachtingen van intelligent agents vrij hoog gespannen zijn. Een paar jaar geleden zocht men de oplossing voor goede zoekresultaten vaak bij agents. Tenslotte worden conclusies getrokken en voorstellen voor de UB gedaan.

⁴ Data mining is het vinden van informatie in grote databanken. Dit kunnen patronen van koopgedrag zijn die belangrijk zijn voor de detailhandel, of afwijkend koopgedrag dat fraude kan betekenen voor een creditcardmaatschappij, e.d. De Nederlandse belastingdienst past dit ook toe op grote aantallen aangiftes.

⁵ Het zoeken en vinden van bepaalde informatie.

⁶ Web directories zijn indexen van onderwerpen die op het Net te vinden zijn.

⁷ Concept inductie is het relateren van begrippen met gelijke of verwante betekenis.

⁸ Startpagina's van waaruit de gebruiker per onderwerp lijsten met Internetadressen kan benaderen.

⁹ Dit zijn programma's die voor een gebruiker naar bepaalde informatie op zoek kan gaan.

2. De gebruiker

De doelgroep van de UB zijn studenten en medewerkers van de EUR. Zij zijn hoofdzakelijk geïnteresseerd in wetenschappelijke informatie. Uit de enquête (zie appendix) bleek dat meer dan 90% van de gebruikers Internet meestal (61%) of soms (31%) gebruikt voor het vinden van wetenschappelijke informatie.

De vraag is wat zoeken die mensen op Internet als slechts 6% van de informatie wetenschappelijk is? Kennelijk genoeg om te gaan zoeken. Anderzijds geeft maar 53% aan dat hun zoektochten min of meer succesvol zijn. Voor de andere 47% die niet tevreden zijn over hun resultaten is het goed hulpmiddelen aan te bieden. Het maakt dan wel uit wat voor hulpmiddelen dat zijn. De meeste startpagina's (www.startpagina.nl), zoekmachines (www.altavista.com) en directories (www.yahoo.com) zijn gericht op het vinden van populaire informatie. Uit de enquête blijkt dat ze toch vaak worden gebruikt.

Startpagina's van bekende universiteitsbibliotheken geven vaak wel goede wegwijzers. Het zou een goede service van de UB zijn ook zo'n startpagina te maken.

Een stap verder is te kijken naar individuele gebruikers. Iedereen heeft een eigen voorstelling van wat er gezocht gaat worden. Ieder begrip (of trefwoord) heeft dus een eigen context, afhankelijk van de voorkennis van de gebruiker, het doel waarvoor hij de informatie nodig heeft, etc. De aangeboden informatie op het Net houdt daar totaal geen rekening mee. Zo hebben boeken een aanduiding als *voor 8 tot 12 jaar* en is wetenschappelijke literatuur door bibliothekmensen geclassificeerd. Voor informatie op het Net kennen we zoiets niet. Het matchen van vraag en antwoord zou een stuk gemakkelijker gaan als we uitgaan van zoiets als mind maps of conceptual schemes¹⁰ zowel aan de kant van de gebruiker als bij de aangeboden informatie.

Zo pleit Humphreys (1999) voor het toekennen van bepaalde metadata¹¹ aan Internetsites. Het automatisch indexeren kan alleen als de site een contextaanduiding geeft, zoiets als *starting and running your own small business*. Zoeken naar het begrip *netwerken* maar dan met *own small business* als context geeft een veel relevanter resultaat dan alleen op *netwerken* zoeken. Zoeken met AskJeeves (www.ask.com) geeft al de mogelijkheid om zo'n mind map te creëren. Jeeves vraagt de gebruiker naar de context: er zijn categorieën om uit te kiezen en er is feedback. In een recent artikel geeft Sieverts (2000) een bespreking van de mogelijkheden van Jeeves. Voor specifieke (wetenschappelijke) informatie werkt deze hulp niet echt goed.

Het gebruik van XML (eXtensible Markup Language) door site makers zal de precisie van de te zoeken informatie vergroten. XML heeft metatags waarin een flink aantal indexeerelementen ondergebracht kunnen worden.

¹⁰ Mind maps en conceptual schemes zijn het beste te omschrijven als trefwoorden met de context die de gebruiker bedoelt.

¹¹ Metadata of metatags zijn de aanduidingen van het soort informatie. *Auteurs* is bijvoorbeeld een metatag bij een of meerdere eigennamen; *Engels* is een metatag bij een Engelstalige bron.

Deze metatags zijn gedefinieerd in de *Dublin Core* afspraken (zie http://purl.org/metadata/dublin_core). Serieuze sites zullen er wel voor zorgen deze metatags correct in te vullen. Net zoals serieuze uitgevers ook zorgen dat hun publicaties de juiste aanduidingen krijgen en zo bij de goede doelgroep terecht komen. Zodra de metatags gemeengoed geworden zijn kunnen we betere informatie van het Internet verwachten.

De metatags kunnen ook behulpzaam zijn bij automatische indexering van sites. Momenteel maken alleen AltaVista, Infoseek en HotBot gebruik van metatags. (Turner en Brackbill 1998)

Bruce (1999) heeft onderzoek gedaan naar cognitieve metaforen die gebruikers hebben als ze naar informatie op het Net gaan zoeken. Zo is er een groep mensen die het Internet beschouwen als een netwerk van wegen, waarbij de dorpen en steden computers voorstellen. Of, ze zien het als een stuk marmer waarbij je de patronen kan volgen. Dit zijn allemaal metaforen die het Net als connectieve structuren opvatten. Anderen zien het Net als een gigantische bibliotheek waar informatie te halen valt.

Nu blijkt dat gebruikers (de test werd gedaan met wetenschappers en studenten) die de laatste analogie gebruiken meer tevreden zijn met hun zoekresultaten dan degenen die het wegensysteem e.d. als metafoor gebruiken. Het lijkt me goed dat de UB hulp over het Net aanbiedt alsof het Net een bibliotheek is.

Vanuit de kant van de informatiespecialisten zijn ook kanttekeningen te maken (Chowdhury 1999). Zo is uit onderzoek gebleken (zie ook hoofdstuk 3) dat het niet duidelijk is hoe en waarom zoekmachines werken. Welke gegevens van de Internetbron hebben ze gebruikt? Omdat de collectie bronnen niet bekend is, is het ook onmogelijk betrouwbare precision en recall¹² gegevens uit te rekenen, hetgeen wel mogelijk is met collecties van conventionele bestanden, zoals cd-rom of online databanken. Men zou een gedeelte van het Net kunnen testen op precision en recall. Waarschijnlijk zal binnenkort TREC¹³ iets dergelijks uitvoeren. Verder zijn Internetbronnen niet gecontroleerd en geëvalueerd door vakgenoten zoals bij conventionele bronnen (boeken, tijdschriftartikelen). Er zijn ook geen standaard user-interfaces waarmee het de gebruiker gemakkelijk gemaakt kan worden. Ovid (www.ovid.com) kent interfaces op conventionele databanken maar zover ik weet niet op Internetbronnen.

En tenslotte is er het al eerder genoemde ontbreken van structuur van het Net, het – meestal – ontbreken van structuur van individuele sites en het ontbreken van classificaties.

Gebruiker en Internet staan nog ver van elkaar. Hulp zal moeten komen van betere technieken die het Internet in termen van de gebruiker te lijf gaan.

¹² Dit is de mate waarin gezochte referenties nauwkeurig en betrouwbaar gevonden kunnen worden.

¹³ TREC (Text Retrieval Conference) biedt een aantal retrieval mogelijkheden voor teksten op Internet, zie http://trec.nist.gov/pubs/trec8/t8_proceedings.html

3 Zoekmachines

3.1 De gewone zoekmachines

De meest bekende manier van zoeken op Internet is het gebruiken van search engines, zoekmachines, zoals AltaVista (www.altavista.com), NorthernLight (www.northernlight.com), en vele andere. In strikte zin zijn het geen zoekmachines maar indexeerprogramma's die voortdurend Internetbronnen bezoeken en de inhoud indexeren. Het zoekgedeelte bestaat uit het vinden van opgegeven trefwoorden in die index en het laten zien van de bijbehorende en ontdubbelde Internet sites.

Onduidelijk is, ook vanwege bedrijfsgeheimen, welke sites nu precies bezocht worden en waarom. Het blijkt ook dat de dekking van de zoekmachines, zelfs in combinatie, maar gering is. Aangezien de meeste zoeksystemen uit de fabriek van Inktomi komen moet geconcludeerd worden dat de zoekfirma's er toch een eigen zoekbeleid op na houden.

Mettrop en Van der Laan (1999) onderzochten dit indexeergedrag en het retrieval gedrag en hun conclusie was dat zoekmachines niet alleen incompetent zijn maar ook onbetrouwbaar. Zij werkten met een aantal documenten, waarvan enige werden aangemeld bij de zoekmachines. Daarna werd met de meest bekende zoekmachines periodiek gezocht. Vreemd was dat documenten die eerder wel gevonden werden later weer onvindbaar bleven en eventueel na verloop van tijd weer terug kwamen. Meer dan de helft van de onderzochte machines vertoont dit inconsistent gedrag. Verder waren er verschillen tussen de zoekmachines voor wat betreft het aantal teruggevonden documenten. De zoekmachines samen vinden niet meer dan tien van de zestien uitgezette documenten. Grote (AltaVista) machines presteerden niet beter dan andere zoekmachines.

De machines die wat betreft retrieval gedrag het best presteerden waren Search.nl, NorthernLight en Lycos. Het slechtst kwamen Vindex, MSN en in mindere mate HotBot uit de test te voorschijn. Voor het volledige onderzoek zie www.cwi.nl/cwi/projects/IRT/.

O'Leary (1999) geeft soortgelijke resultaten en signaleert tevens allerlei fouten die op webpagina's voorkomen en daardoor verkeerde zoekresultaten opleveren.

De conclusie is duidelijk: voor een *fast and dirty* zoektocht over Internet zijn de bekende zoekmachines niet slecht en leveren wellicht in eerste instantie wat interessante ingangen. Verwacht echter niet dat er goed en gedegen gezocht is. Verwacht beslist geen kwaliteitsdocumenten, die komen slechts toevallig naar boven tussen een heleboel rommel.

Het is echter niet altijd onverstandig zoekmachines te gebruiken. In gevallen dat heel gericht gezocht kan worden op termen die niet al te veel voorkomen is het goed een lijstje van bekende zoekmachines paraat te hebben. Een recent overzicht is te vinden op www.searchenginewatch.com. Hierin staan ook de

metacrawlers die combinaties van de resultaten van de gebruikelijke zoekmachines opleveren.

Belangrijk is dan ook te weten hoe gezocht kan worden. Soms kan Booleaans (met AND en OR), soms met wild cards (bij twijfel aan de spelling) gezocht worden en verder met al de trucs die informatiespecialisten wel kennen. Internetgebruikers kennen die trucs vaak niet. Daarom zijn overzichten met *wat kan en hoe het moet* goede begeleiders voor de gemiddelde Internetgebruiker.

3.2 Portals

Bij de zoekmachines is nog een categorie te onderscheiden, de zogenoemde portals, directories of gidsen. Er wordt eerst naar een categorie gevraagd (*travel* of *people* en dat soort indelingen), binnen een categorie kan weer verder geselecteerd worden en uiteindelijk gaat de zoekmachine op zoek in zijn index. Meestal zijn deze gidsen handmatig gemaakt. Mijn indruk is dat ze niet echt op de wetenschappelijke onderzoeker zijn gericht. De bekende zijn Yahoo! (www.yahoo.com) en Argus Clearinghouse (www.clearinghouse.net) die ook nieuwsgroepen verzamelt.

4. De betere zoekmachines

Met de onvolwaardigheid van de bekende zoekmachines hoeven we niet verder te leven nu er een nieuwe generatie wordt ontwikkeld. Op het moment zijn er twee producten die waarschijnlijk de trend zullen zetten voor toekomstige zoekmachines, de searchbots¹⁴. De ene is Google (al operationeel: www.Google.com) en de andere Clever (werkt nog niet, meer gegevens zie www.almaden.ibm.com/cs/k53/clever.html¹⁵). Google is ontwikkeld door mensen van Stanford University, Clever door IBM. De naam Google is afgeleid van het woord googol (1 met 100 nullen). Op dit moment heeft Google 560 miljoen pagina's geïndexeerd en 500 miljoen gedeeltelijk geïndexeerd.¹⁶ Dat is veel voor een zoekmachine, bovendien is het kwalitatief in orde. Het leuke van Google is ook dat er vrijwel geen reclames te zien zijn, behalve één reclameberichtje dat bovendien betrekking heeft op de zoekterm.

Het idee voor zowel Google als Clever is dat er twee soorten bruikbare sites zijn: de autoriteiten en de hubs. Eigenlijk is dit een toepassing van het herkennen van sociale netwerken en citatie-indexen. Autoriteiten zijn sites waarvan we weten dat ze goed zijn, bijvoorbeeld homepages van grote organisaties en universiteiten. Autoriteiten zijn ook te vinden omdat veel andere sites naar hen verwijzen. Die verwijzende sites zijn hubs of springplanken. Hubs die veel naar autoriteiten

¹⁴ "Searchbot" is de samenstelling van search engine en robot; m.a.w.: een intelligente zoekmachine.

¹⁵ Na 1999 is niet veel meer over Clever vernomen. IBM geeft geen response op e-mails hierover.

¹⁶ Inmiddels (zomer 2001) zijn er 1,75 miljard pagina's door Google geïndexeerd.

verwijzen zijn goede hubs. IBM heeft algoritmen bedacht die gebruikmakend van deze hub-autoriteit verhouding DE goede autoriteiten op een bepaald gebied aanwijzen. Het verzamelen van autoriteiten op een bepaald gebied is een bijproduct van het Clever project (ARC = automatic resource compilation) zie <http://www7.scu.edu.au/programme/fullpapers/1898/com1898.html>) en www.almaden.ibm.com/almaden/feat/www8. Dit moet leiden tot het maken van echt goede subject gateways (zie volgend hoofdstuk) zonder menselijke inspanning, up-to-date en hopelijk volledig. Dit wordt focused crawling genoemd. Bij Google kunnen ook gerelateerde sites worden bekeken, bijvoorbeeld zoek op Stanford en met GoogleScout kun je naar Harvard University, etc. komen. Google heeft kopieën van alle geïndexeerde pagina's. Als een server van een te bezoeken site uit de lucht is, of een bladzijde tijdelijk verdwenen, zorgt Google toch nog voor een kopie. Tenslotte kan men via Google ook betaalde zoekopdrachten geven. Zie www.google.com/doing_business.html.

5. Subject Gateways

5.1 Inleiding

Het is uit vele publicaties en eigen ervaring duidelijk dat het zoeken met de traditionele zoekmachines zoals AltaVista, NorthernLight e.d., een mer à boire oplevert. De superzoekmachines of metacrawlers maken de hoeveelheid informatie alleen maar groter. Waar de wetenschapper behoefte aan heeft is een overzichtelijke hoeveelheid informatie die bovendien kwaliteit bezit. Dit kwam ook naar voren bij de uitkomst van de enquête. Een flink aantal inzenders vroeg naar bronnenlijstjes. Het is prettig te weten dat er al mensen bezig zijn geweest met het zoekwerk en de kwaliteitscontrole. Al vanaf de prille dagen van Internet toen er nog geen World Wide Web was, maar Gophers en Veronica's, zijn er wetenschappers c.q. hobbyisten geweest die het Internet afstruinden naar goede sites op hun vakgebied en die – als waren het vlinders – verzamelden in hun eigen lijstjes. Deze lijstjes gingen een eigen leven leiden op homepages van faculteiten; er werd over en weer gepikt uit andere lijstjes en zo ontstonden de onderwerpsdirectories. Deze directories, ook wel gateways genoemd, hebben talloze voordelen. Ten eerste bieden ze een lijst met sites die geselecteerd zijn door wetenschappers uit de eigen discipline. Ten tweede zijn ze enigszins overzichtelijk en bovendien zijn er meestal zoekprogramma's gemaakt, die binnen de hoeveelheid sites nog eens op trefwoorden of namen kunnen zoeken. De opbrengst is dan behapbaar en de moeite waard om één voor één te bekijken. Ten derde kan men er redelijkerwijs vanuit gaan dat er niet veel gemist is. De gezamenlijke wetenschappelijke Internetzoekers gaan er prat op hun werk goed te doen (zie Beekink 1999 en Hofstede 1998). Hofstede heeft een grote collectie van deze directories verzameld: www.leidenuniv.nl/ub/biv/specials.htm.

Beekink noemt in zijn artikel talloze wetenschappelijke directories met hun beschrijvingen.

Er zijn nu ook tal van organisaties die subject gateways samenstellen op een professionele manier. De professionele gateways zijn verzamelingen Internetbronnen gerubriceerd naar onderwerp. Ze worden ook wel gidsen of directories genoemd, niet te verwarren met de gewone zoekmachines die er categorieën op na houden. De filosofie die aan de gateways ten grondslag ligt, is het structureren van de chaos die Internet heet.

Subject gateways zijn allemaal gebaseerd op hetzelfde principe: onderwerp- en informatiespecialisten selecteren bronnen, beoordelen ze op bruikbaarheid, geven er beschrijvingen van en kennen trefwoorden toe. Daarna worden deze bronnen regelmatig gecontroleerd of ze nog bestaan en of de inhoud niet te veel is veranderd.

Een goed voorbeeld is de ERIC organisatie die al jaren de databanken voor educatief materiaal samenstelt. ERIC verzamelt nu ook Internetbronnen samen in een consortium van momenteel (juli 2000) 100 andere organisaties en bibliotheken op educatief gebied. Deze Internetverzameling heet GEM, the Gateway to Educational Materials (www.geminfo.org), die er zeer indrukwekkend uitziet.

5.2 DESIRE

DESIRE is een Europees initiatief om een verzameling subject gateways op te bouwen. (Worsfold 1998) De kracht van zo'n werkwijze is dat er wereldwijd wordt verzameld en dat de bronnen op één URL te vinden zijn.

Er zijn verscheidene (ik vermoed: heel veel) initiatieven op dit gebied. Ze zijn allemaal gebaseerd op dezelfde uitgangspunten: afspraken over de reikwijdte (zoals vakgebied en geografische scope), er zijn kwaliteitscriteria en er zijn afspraken over de wijze van verzamelen en van onderhoud. Zo is het soms ook mogelijk om voorstellen voor nieuwe bronnen aan te dragen. De onderwerpspecialisten beoordelen deze en geven ze een plaats in de virtuele collectie.

De opzet van DESIRE is het creëren van een netwerk van specialisten die ieder hun bijdragen aanleveren op een standaard manier. Het samenwerken resulteert in vermijden van doublures en dus van onnodig extra werk. De gebruiker profiteert van kwalitatief hoogwaardige bronnen die gerubriceerd naar onderwerp zijn en bovendien een beschrijving hebben.

In Nederland is DutchESS de subject gateway. Het participeert in het DESIRE-project, maar verzamelt alleen wat door een handjevol vakreferenten in Nederland wordt aangedragen. Het mist een aantal van de voordelen die wereldwijd opereren biedt. Zo is het aanbod erg beperkt in vergelijking met wat er al op het onderwerp is verzameld. Als de aanvoer van nieuwe bronnen wereldwijd gecoördineerd zou geschieden, zou het een waardevolle onderneming zijn om hieraan deel te nemen. Of, als DutchESS zich had toegelegd om als enige ter wereld Nederlands materiaal (dus over Nederlandse onderwerpen, eventueel in

andere talen) te verzamelen, was het een product met meerwaarde geweest. Nu is het een betrekkelijk marginaal gebeuren. Een ander nadeel is de versluiering van de bronnen. Door het maken van abstracts en het toekennen van trefwoorden en NBC-codes¹⁷ wordt een bibliotheektechniek losgelaten op een medium dat daar niet voor bedoeld is. Het vinden van specifiek materiaal is juist gemakkelijker door de full text doorzoekbaarheid van een Internetsite. Bij DutchESS is dat nog altijd niet mogelijk. Het zou voor vakreferenten veel waardevoller zijn mee te werken aan subject gateways die op veel grotere schaal opereren.

5.3 Evaluatie

De voordelen van de subject gateways boven het zoeken met zoekmachines zijn de moeite waard:

1. De bronnen staan per vakgebied bij elkaar, meestal d.m.v. een bekende classificatiemethode.
2. Doordat er niet noodzakelijkerwijs een trefwoord hoeft te worden opgegeven is het mogelijk door een aantal bronnen te browsen.
3. De bronnen zijn verzameld door vakgenoten op grond van kwaliteitscriteria. Dit betekent dat er voornamelijk hoogwaardige wetenschappelijke bronnen bij elkaar staan.
4. De bronnen zijn voorzien van metadata, d.w.z. dat ze meestal een beschrijving hebben, informatie over de soort bron en trefwoorden. De beschrijving kan voor de gebruiker een indicatie zijn om de bron al dan niet zelf te openen.
5. Vermelding van nieuwe bronnen op het Net. Zo hebben bijvoorbeeld de gateway voor sociale wetenschappen SOSIG en bovengenoemde GEM een *wat is nieuw*-knop. Hiermee kan de gebruiker dagelijks of wekelijks kijken wat er nieuw is bijgekomen.

Eigenlijk zijn deze gateways prachtige instrumenten. Toch moeten er wat kanttekeningen bij geplaatst worden:

1. Er is een discrepantie tussen de opvattingen van kwaliteit en/of gebruikswaarde bij de informatiespecialist en de eindgebruiker. Een site met plaatjes en anekdotes over een bepaald onderwerp kan interessant zijn voor een docent die zijn colleges wat wil opfleuren. De specialist neemt zo'n site niet op omdat deze niet aan de kwaliteitscriteria voldoet.
2. Het probleem van versluiering. Het indelen in categorieën is voor de gebruiker niet altijd eenduidig. Gebruikers werken graag met trefwoorden en dat gaat soms mis. Het zoeken op trefwoord in DutchESS bijvoorbeeld slaat niet op de complete teksten van de bronnen maar op de annotaties van de indeler. Alle bronnen met een soort verzamelkarakter (lijstjes, encyclopedieën) kunnen niet met specifieke trefwoorden beschreven worden. Dit probleem is op te lossen door de bronnenverzamelingen met een search

¹⁷ Nederlandse Basis Classificatie

engine te indexeren, zoals bij de meeste andere professionele verzamelingen wel het geval is.

3. De bronnen zijn niet altijd vers van de pers. Er kan een aanzienlijke tijd voorbij gaan tussen publicatie op het Net en het toevoegen aan de verzameling.
4. Voor de gebruiker is niet duidelijk welke de criteria voor opname zijn en hoe de dekking van het totale aanbod is. Een verantwoording van de gateway-organisatie zou onder een knop opvraagbaar moeten zijn. Dit is overigens ook het geval voor het zoeken met zoekmachines.
5. Voor de gebruiker is het heel vervelend dat er zoveel verschillende internationale initiatieven zijn waardoor een nieuwe chaos is geschapen.

5.4 Een willekeurig lijstje van subject gateways:

ADAM (kunst, design, architectuur, media)

<http://www.adam.ac.uk/>

Biz/ed (bedrijfskunde en economie)

<http://bized.ac.uk/>

GEM (onderwijs en opvoeding)

www.geminfo.org

Geschiedenis

<http://ihr.sas.ac.uk/>

Lokhorst's philosophy sites (filosofie)

<http://www.eur.nl/fw/staff/lokhorst/links.html>

MathGuide (wiskunde)

<http://www.sub.uni-goettingen.de/ssgfi/math/index.html>

OMNI (geneeskunde, gezondheidsmanagement)

<http://omni.ac.uk>

SOCIOSITE (sociale wetenschappen)

www.pscw.uva.nl/sociosite/index.html

SOSIG (sociale wetenschappen, incl. economie, rechten en filosofie)

<http://www.sosig.ac.uk/>

BUBL Link (algemeen)

<http://bubl.ac.uk/link/>

DutchESS (algemeen)

<http://www.kb.nl/dutchess/>

NetFirst (algemeen; alleen voor abonnees)

<http://www.oclc.org/oclc/netfirst.htm>

NISS (algemeen)

<http://www.niss.ac.uk/subject/index.html>

Verder zijn er mooie verzamelingen te vinden via grote universiteitsbibliotheken. Een goed voorbeeld is te vinden bij de UB van Utrecht (www.library.uu.nl). Zie verder de reeds eerder genoemde verzameling op

www.leidenuniv.nl/ub/biv/specials.htm.

Het zelf gaan verzamelen van bronnen door de UB, eventueel aangedragen door de faculteiten, zou een erg arbeidsintensieve activiteit zijn en bovendien wordt er elders al zoveel verzameld. Het meest efficiënte is om bestaande verzamelingen aan te bieden, eventueel met specifieke wensen uit de faculteiten (zie enquête wensen).

6 Intelligent agents

6.1 Inleiding

Kunstmatige intelligentie, in deze context, wordt omschreven als een systeem dat op grond van informatie en bepaalde regels kennis kan voortbrengen. Robots kunnen op grond van informatie uit de omgeving en volgens bepaalde regels zinvol (intelligent) gedrag vertonen. Kennissystemen en robots werken omdat hun *universe of discourse* vrij beperkt is. Goede voorbeelden zijn schaakcomputers en puntlassende robots. Zodra de belevingswereld of omgeving te groot wordt, zoals het Net, werken deze systemen niet meer. Intelligente systemen zijn ook vaak uitgerust met heuristiek, d.w.z., regels die in geval van twijfel kunnen worden ingezet. Zo heeft een schaakcomputer heuristieken om te weten welke positie de beste kans van slagen heeft en de moeite waard is om een aantal zetten vooruit uit te rekenen. Zo zou een intelligent agent op het Net moeten weten wanneer hyperlinks in de breedte (allemaal, maar dan één laag diep) of in de diepte (eentje, maar dan telkens een laag dieper) bezocht moeten worden. Zo'n beslissing zou een mens maken op grond van de plaats van een link in de tekst. Helaas kunnen intelligent agents geen natuurlijke taal begrijpen, dus ook niet dat soort beslissingen nemen. Wel is het mogelijk om op een beperkt terrein, bijvoorbeeld alleen in nieuwsberichten, taal te begrijpen. Hier zien we dan ook dat intelligent agents inzetbaar zijn. Voor het zoeken naar wetenschappelijke informatie is zoiets vooralsnog heel lastig. Een alternatief is dan het lerende systeem. Deze systemen leren door voorbeelden (acceptatiesystemen bij verzekeringsmaatschappijen), door analogie (*meer zoals dit*), of door feedback (bijsturing door de gebruiker).

Voor intelligent agents die voor ons op het Net gaan zoeken zijn eigenlijk alleen lerende systemen van toepassing.

Intelligent agents worden ook wel software agents, knowbots of softbots genoemd. Een intelligent agent is in de meest eenvoudige vorm een SDI¹⁸: op regelmatige tijden wordt een aantal zoekmachines tegelijk (metasearch) geraadpleegd en de resultaten worden automatisch naar de gebruiker gezonden. Dit heeft eigenlijk niets met intelligent gedrag te maken. Zo is My Yahoo! (www.yahoo.com) een service waarbij een interessegebied kan worden opgegeven. Er wordt voor de gebruiker een eigen pagina gecreëerd die hij van tijd tot tijd kan bezoeken om het laatste nieuws op zijn interessegebied te lezen. Dat is niet echt intelligent. Lycos, Excite en HotBot hebben deze service ook. Een andere service is die van Alexa (www.alexa.com). Deze levert achtergrondinformatie over de organisatie die een gevonden site publiceert, zodat de gebruiker de betrouwbaarheid van de informatie beter kan beoordelen. Alexa heeft ook een webarchief zodat gekeken kan worden hoe gegevens van bepaalde sites veranderen; of om naar *verdwenen* pagina's met interessant materiaal te zoeken. (Bradley 1999)

De betere intelligent agents zijn zoekmachines die een lerend vermogen hebben. Uiteindelijk weten de agents de beste links te vinden. Dit kan worden versterkt door feedback naar de gebruiker.

Adaptive agents kunnen ook rekening houden met de context waarin ze zoeken. Dit geeft ook weer een voorsprong op de gewone zoekmachines.

Een speciale vorm hebben de multi-user agents die sites aanbevelen die ook bezocht worden door andere gebruikers met hetzelfde belangstellingsprofiel.

Een probleem blijft dat de intelligent agent gevoed wordt door een gebruiker die vaak niet precies weet hoe trefwoorden gebruikt moeten worden. Wetenschappers die juist innovatieve ontwikkelingen willen volgen hebben ook niet het goede vocabulaire om de juiste trefwoorden te geven. Ze weten immers niet exact waar ze naar zoeken.

Hieronder volgen een paar voorbeelden van intelligent agents, waarbij alleen InfoSpiders en de shopping agents min of meer succesvol zijn. Echte aanbevelingen zijn er niet bij.

6.2 InfoSpiders

Een voorbeeld van een autonome (zonder hulp van de gebruiker) agent is InfoSpiders. Deze werkt op bestaande URL's van sites die voor de gebruiker *goed* zijn maar nog niet echt wat ze zoeken. Uitgaande van zo'n site wordt InfoSpiders geïnitieerd en zoekt in de omgeving (links) van de eerder gevonden

¹⁸ Selective Dissimination of Information

sites. Uiteindelijk komt een lijst met naar relevantie gesorteerde sites terug. Dit is wel een tijdrovende procedure. (Menczer en Monge 1999)

6.3 Kenjin

The Autonomy Corp. is al een tijdje bezig met het ontwikkelen van intelligent agents. In het artikel van Tegenbos 1997 wordt de agent *Autonomy* besproken. Het werkt met natuurlijke-taal vragen, voert uit zichzelf zoekacties uit en doet iets aan de aanbieder van de verzamelde informatie. Hoe het precies werkt is vooralsnog een bedrijfsgeheim. De resultaten zijn echter niet imponerend. Ten eerste gebruikt Autonomy dezelfde zoekmachines die, zoals we in een vorig hoofdstuk zagen, niet geweldig presteren. Het leermechanisme van Autonomy berust op het vragen naar *meer van dit soort sites?* maar dat kan ook totaal een verkeerde response opleveren, omdat de gebruiker vaak een andere reden heeft om een site te waarderen dan de intelligent agent, die eerder op de waarschijnlijkheid van veel voorkomende begrippen let. Een recente opvolger van Autonomy bij dezelfde organisatie is Kenjin (www.kenjin.com). Kenjin (Japans voor wijze man) is een programma dat via deze site gedownload kan worden. Op je eigen PC houdt het alles in de gaten. Het analyseert alle documenten die voorbijgaan, zoals e-mail, zoeken op Intra- en Internet, en zelfs Worddocumenten. Op grond van deze analyses voert Kenjin zoekacties uit op het Net maar ook op je eigen documenten. De release is vrij recent (zie Time, April 10, 2000 en Intermediair, 13 april 2000).

6.4 Shopping agents

Een speciale groep van intelligent agents zijn de shopping agents. In de begin-dagen van commerciële sites op Internet waren deze al actief met het zoeken naar concurrentie. Zo had Hertz een agent die elke dag de prijzen van huurauto's op de grote vliegvelden afgraasde, met de bedoeling een nog scherpere prijs per vliegveld te kunnen aanbieden.

Tegenwoordig worden er ook shopping agents via Internet aangeboden voor alle gebruikers (Jasco 1999). Shopping agents zoeken naar allerlei koopwaar op het Internet voor de laagste prijs. Allerlei variaties zijn hierbij mogelijk. In het artikel van Jasco wordt een overzicht gegeven van voor- en nadelen van de verschillende shopping agents. Zo kwam de shopping agent van Alta Vista (www.shopping.com) er niet gunstig uit wat betreft het vinden van de goedkoopste aanbieding. Alta Vista heeft wel een extraatje en dat is een grote verzameling van stukjes muziek van 30 seconden om de cd's te kunnen beoordelen. Yahoo's shopping agent (www.shopping.yahoo.com) levert meer op dan waar om gevraagd wordt. Bottom Dollar (www.bottomdollar.com) is meer precies en levert koopjes op. Belangrijk is dat Bottom Dollar filters gebruikt om specificaties aan te geven, bij muziek kan men bijvoorbeeld aangeven of het om een cd, cassette of lp gaat.

Concluderend kan men zeggen dat de technologie van dit soort agents goed werkt op een duidelijk gedefinieerd gebied. De zoekvraag kan vrij precies geformuleerd worden en de te zoeken sites zijn de bekende commerciële e-winkels die zich goed profileren en daardoor bij de shopping agents bekend zijn.

7. Resultaten

7.1 Conclusie

Uit de voorgaande hoofdstukken zijn de volgende onderwerpen naar voren gekomen.

- Gebruikers hebben over het algemeen niet goed grip op wat Internet te bieden heeft. Als er een bibliotheekachtige omgeving voor ze gecreëerd wordt d.m.v. een goede startpagina waar ze zowel met trefwoorden als met onderwerpslijstjes verder kunnen, zou dat meer vertrouwdheid met het Net ten goede komen. Waarschijnlijk zullen hun zoekresultaten ook verbeteren, wat weer vertrouwen in het Net opwekt, enzovoorts.
- Zoekmachines zijn niet echt de beste hulpmiddelen om te zoeken naar wetenschappelijke informatie.
- Er zijn wel ontwikkelingen om de zoekmachine beter te maken. Zo zijn Google (www.google.com) en Clever (nog niet operationeel) methodes om te zoeken naar sites die al een status van betrouwbaarheid en/of aanbevelenswaardigheid van een bepaalde gemeenschap hebben opgebouwd. De relevantie en kwaliteit van de gevonden sites is veel groter geworden.
- Deze of andere zoekmachines krijgen flink wat meerwaarde als voor de gebruiker duidelijk is op wat voor manier gezocht kan worden.
- Subject gateways kunnen voor gebruikers op eenvoudige wijze een schat aan informatie bieden. Er is echter een gigantisch aanbod zodat de gebruiker alsnog in het bos verdwaalt. Tot nu toe ben ik geen wereldwijde consensus over de beste gateways tegengekomen. Kijken bij de grote universiteitsbibliotheken leverde steeds andere lijstjes op.
- Subject gateways zijn alleen goede gidsen als ze per gateway door een full text zoekmachine doorzocht kunnen worden. Anders (zoals bij DutchESS) zijn ze slechts aanreikingen voor browsen om zo toevallig iets te vinden. Bij grote bronnencollecties is browsen alleen al een tijdrovende aangelegenheid.
- Het succes van intelligent agents is slechts te vinden bij specifieke onderwerpen en dan nog niet erg overtuigend. De verwachtingen van vier of vijf jaar geleden zijn nog niet waargemaakt. Dit kan nog komen wanneer er beter gebruik van metatags wordt gemaakt zowel door de sitemaker als door de zoekmachine van de intelligent agent.
- Voor betere matches van vraag en antwoord kunnen mind maps en conceptual maps uitkomst bieden. Hiervoor geldt echter ook dat ze nog niet goed werken.

7.2 Uitkomsten naar aanleiding van de vraagstelling

In een vooronderzoek werden een vijftal vragen geformuleerd. In deze paragraaf wordt gekeken in hoeverre antwoorden hierop gegeven kunnen worden.

1. Wetenschappelijk versus commerciële en niet-wetenschappelijke informatie.

Uit de enquête kwam naar voren dat 61% meestal en 31% soms wetenschappelijke informatie zoekt; bij FSW is dit 50% meestal en 50% soms. Een aantal keren kwam de wens naar voren dat het prettig zou zijn alleen op wetenschappelijke informatie te kunnen zoeken.

Intelligent agents die alleen gericht zijn op wetenschappelijk onderzoek bestaan nog niet. Wel kan gewezen worden op zoekmachines die het kaf van het koren kunnen scheiden zoals Google. Subject gateways geven ook kwaliteitsgarantie van de bronnen. Een apart wetenschappelijk Net zou ook tegemoet kunnen komen aan de wens om uitsluitend wetenschappelijke informatie te vinden.

In de inleiding gaf ik al aan dat men zich niet moet blind staren op de mogelijkheden van Internet. Veel meer wetenschappelijke informatie is te halen uit de gerenommeerde databanken met gerecenseerd materiaal.

2. Zoeksystemen versus intelligent agents.

Deze vraag kan op dezelfde wijze beantwoord worden. Intelligent agents kunnen alleen op afgebakende gebieden iets bieden. De betere zoeksystemen zijn alleszins acceptabel.

3. Informatie op het Internet versus informatie die men nodig heeft.

Uit de enquête bleek dat 47% van de zoekenden meestal niet tevreden is met de zoekresultaten. Aanbieden van een goede startpagina met de betere zoekmachines inclusief de hierbij gegeven tips, alsmede een lijst van gateways op onderwerp en wellicht ook een cursus, zou al veel helpen. Het volgen van de ontwikkeling van zoekbots die op een intelligente manier vragen vertalen is een andere aanrader.

4. Wetenschappelijke informatie in het algemeen (chaos) versus vakgerichte informatie (structuur).

In deze stelling worden ook weer de subject gateways genoemd. Een aantal van degenen die de enquête beantwoord hebben, noemde het beginnen met een lijstje bronnen een goede zoekmethode. Ook waren er vragen naar zulke lijsten. Zie verder de opmerkingen onder punt 1.

5. Gratis versus niet-gratis informatie.

Zie hiervoor ook de reactie onder punt 1. Bibliografische databanken zijn meestal campusbreed gratis te raadplegen. Uit de enquête bleek dat het verschil tussen gratis Internet en dure, maar voor de gebruiker gratis *conventionele* databanken voor die gebruiker niet altijd duidelijk is.

Tijdens het onderzoek is getracht ontwikkelingen te vinden op het gebied van systemen die het zoeken op Internet vergemakkelijken. Dit bleek een onmogelijke opgave omdat met dit soort systemen veel toekomstig geld op het spel staat en ze niet zomaar op het Net aan de openbaarheid worden prijsgegeven. Toevallige contacten met een ontwikkelaar moeten hiervoor uitkomst brengen.

7.3 Voorstellen voor een UB-helppagina voor Internet

Voortbordurend op de conclusies uit 7.1 en de uitkomst van de enquête (zie appendix) zou ik voor de UB/EUR aanbevelen een startpagina te ontwerpen waarop een lijstje met goede zoekmachines en hun specifiek gebruik is te vinden. Ik denk dan aan een lijstje met subject gateways per vakgebied (en full text retrieval!), een pagina met Internettips en/of een cursus en eventueel een knop naar een helpdesk van de UB. In de bovenbalk van de UB-homepage zou een knop *Internetgids en tips* moeten zitten, waarachter zich de startpagina bevindt. Uiteraard moeten de gegevens regelmatig worden gecontroleerd en aangevuld.

Een voorbeeld van zo'n pagina is die van de universiteitsbibliotheek van Berkeley: <http://www.lib.berkeley.edu/Help/search.html> .

Een voorbeeld van een uitgebreide leidraad, in dit geval voor business researchers, is te vinden op <http://www.brint.com/userguide.htm> .

Een algemene leidraad die achter een knop *meer over Internet* zou kunnen staan, is te vinden bij <http://www.searchenginewatch.com/facts/index.html>. Deze pagina is ontwikkeld door Stanford University.

Een cursus Internetgebruik met veel verwijzingen naar bronnen, ontwikkeld door de KUB, is online te volgen op <http://cwis.kub.nl/~dbi/instruct/www/index.htm>

Een cursus met algemene zoektechnieken, dus niet alleen voor Internet, wordt aangeboden door de Universiteit Twente. Een demo is te vinden op: http://www.ub.utwente.nl/diensten/meewiz/MEEWIZ_nl.shtml

8. Referenties

Beekink 1999. Marcel Beekink. Een andere benadering van het Web. *In: Informatie Professional. no 9 1999*

Bradley 1999. Phil Bradley. Intelligent agents on the Web. *In: Managing Information. Jan/Febr. 1999*

Bruce 1999. Harry Bruce. Perceptions of the Internet: what people think when they search the Internet for information. *In: Internet research: electronic networking applications and policy. no 3 1999*

Chowdhury 1999. G.G. Chowdhury. The Internet and information retrieval research: a brief overview. *In: Journal of Documentation. March 1999*

Hofstede 1998. Marten Hofstede. Speciale zoekmachines op Internet. *In: Informatie Professional. no 12 1998*

Humphreys 1999. Nancy K. Humphreys. Mind maps: hot new tools proposed for cyberspace librarians. *In: Searcher. no 6 1999*
www.infotoday.com/searcher/jun/humphreys.htm

Intelligent agents in cyberspace 1999. Papers from the 1999 AAAI Spring symposium. Technical Report SS-99-03

Jasco 1999. Peter Jasco. New Web technology: Shopping agents. *In: Information Today. Dec.1999*

Klusch 1999. Matthias Klusch, ed. Intelligent information agents; agent-based information discovery and management on the Internet. (Springer 1999)

Lawrence en Gilles 1999. Steve Lawrence en C. Lee Gilles. Accessibility of information on the web. *In: Nature. 8 July 1999*

Menczer 1999. Filippo Menczer en Alvaro E. Monge. Scalable Web search by adaptive online agents; an InfoSpiders case study. *In: Klusch 1999.*

Mettrop 1999. Wouter Mettrop en Hans van der Laan. WWW-zoekmachines onderzocht. *In: Informatie Professional. no 3 1999*

O'Leary 1999. Daniel E. O'Leary. Internet-based information and retrieval systems. *In: Decision support systems. 1999*

Sieverts 2000. E.G. Sieverts. Ask Jeeves – Je kunt hem alles vragen, maar of hij je altijd begrijpt.... *In: Informatie Professional. Juli/augustus 2000*

Tegenbos 1997. J. Tegebos en P. Nieuwenhuysen. My kingdom for an agent? Evaluation of Autonomy, an intelligent search agent for the Internet. *In: Online & CD Rom review. June 1997*

Turner 1998. T.P. Turner en L. Brackbill. Rising to the top: evaluating the use of the HTML META tag to improve retrieval of World Wide Web documents through Internet search engines. *In: Library resources & technical services. Oct 1998*

Worsfold 1998. E. Worsfold. Subject gateways: fulfilling the DESIRE for knowledge. *In: Computer networks and ISDN systems. 1998*

APPENDIX Uitslag enquête

In maart 2000 werd een online-enquête verstuurd aan het gehele wetenschappelijke personeel van de zeven faculteiten op Woudestein. Het formulier is te vinden op (www.eur.nl/ub/FORMS/enquete.htm). Er waren precies 250 respondenten. Dubbele exemplaren (per ongeluk tweemaal verzonden) zijn verwijderd. Eén of twee respondenten waren waarschijnlijk niet serieus.

De enquête was op drie aandachtspunten gericht. De opzet van de enquête was om in een beknopte vraagstelling (vijf minuten invultijd) gegevens boven water te krijgen die voor de vraagstelling van het onderzoek van belang zijn. Ten eerste de vraag naar gebruikfrequentie en de beoordeling van de resultaten. Ten tweede vragen over het soort informatie dat gezocht wordt en de wijze waarop. Ten derde twee vragen die met vrije tekst beantwoord kunnen worden over tips en wensen over gebruikte technieken.

1. Zoekfrequentie en beoordeling van de zoekresultaten.

Gevraagd werd hoe vaak op Internet gezocht werd (vraag 1).

- vaak (> 1 uur per dag)
- geregeld (> 1 uur per week)
- zelden of nooit.

Hierna werd gevraagd de resultaten als volgt te beoordelen:

- *goed en bruikbaar* (vraag 2)
- *te veel* (vraag 3)
- *te veel niet relevant* (vraag 4)
- *te weinig of niets* (vraag 5)

Voor iedere beoordeling kon *meestal*, *soms* of *zelden of nooit* aangegeven worden.

Hieruit zijn succesvolle en niet-succesvolle zoekers te componeren.

De succesvolle zoekers zijn degenen die de resultaten **meestal goed en bruikbaar** vinden, in combinatie met antwoorden **soms** of **zelden of nooit** op de vragen 3, 4 en 5.

De niet-succesvolle zoekers geven voor de beoordeling *goed en bruikbaar* **soms** of **zelden of nooit** in combinatie met antwoord **meestal** of **soms** voor de vragen 3, 4 en 5.

Overigens kunnen deze zoekers ook minder succesvol genoemd worden. Zo zwart-wit is de onderscheiding niet.

Hieronder volgt een lijstje van de scores vergeleken met de frequentie van het Internetgebruik per faculteit (met tussen haakjes de aantallen per faculteit).

FEW = Faculteit der Economische Wetenschappen
 FRG = Faculteit der Rechtsgeleerdheid
 FBK = Faculteit der Bedrijfskunde
 FSW = Faculteit der Sociale Wetenschappen
 BMG = Instituut Beleid en Management Gezondheidszorg
 FHKW = Faculteit der Historische en Kunstwetenschappen
 FWB = Faculteit der Wijsbegeerte

	zoekt vaak	zoekt vaak	zoekt geregeld	zoekt geregeld	zoekt zelden of nooit	zoekt zelden of nooit
	Succesvol	n. succesv.	succesvol	n.succesv.	succesvol	n.succesv
FEW (104)	14	9	34	39	2	6
FRG (40)	7	6	13	12	1	1
FBK (35)	10	2	12	6	1	4
FSW (30)	4	0	9	15	0	2
BMG (19)	6	2	5	4	1	1
FHKW (16)	5	1	4	6	0	0
FWB (6)	3	1	1	1	0	0
TOTAAL (250)	49	21	78	83	5	14

Tabel 1. Zoekfrequentie en succesvol of niet en minder succesvol zoeken.

Het is niet verwonderlijk dat er tweemaal zoveel succesvolle zoekers zijn bij degenen die meer dan een uur per dag zoeken, tegenover 50/50 bij de geregelde zoeker (een uur of meer per week) en driemaal zoveel niet of minder succesvolle zoekers bij de zelden-of-nooit-zoekers. Het is wel verontrustend dat bij de frequente zoekers 30% minder of niet succesvol zijn. Deze mensen verdoen dus elke dag minstens een uur met slecht zoeken. De geregelde zoekers zijn voor de helft niet of minder succesvol, dat is ook zonde van de tijd. Het is wel zo dat niet iedereen dezelfde definitie van *goed en bruikbaar*, e.d. heeft. Door de grote aantallen (van alle inzenders: 132 wel succesvol en 118 niet of minder succesvol) kan men echter stellen dat bijna de helft van de Internetgebruikers op de EUR niet het gewenste resultaat van zoeken op Internet bereikt.

2. De tweede cluster vragen was bedoeld om meer inhoudelijke aspecten naar boven te halen.

Vraag 6 luidde *gebruikt u wetenschappelijke informatie (in tegenstelling tot commerciële informatie)?*

De meeste respondenten antwoordden: 61% meestal, 31% soms en 8% zelden of nooit. Alleen bij de FSW lag de verhouding 50% meestal, 50% soms en 0% zelden of nooit. Wellicht ligt deze afwijking aan het lage aantal (n = 30), of gebruiken de sociologen meer niet-wetenschappelijke informatie voor hun werk. In de EUR wordt veel niet-wetenschappelijke informatie gezocht, hoewel het een academische instelling is. Een volgende vraag zou zijn hoe deze niet-wetenschappelijke informatie in het wetenschappelijke werk past en of deze informatie soms alleen voor hobby of persoonlijke interesse bedoeld is. Daar staat tegenover dat 92% van de zoektochten *meestal* en 100% van de zoektochten *soms* wetenschappelijk is. Ik noem deze afweging omdat er sprake is het Net op te delen in een zuiver wetenschappelijk netwerk en de rest.

	meestal	Soms	zelden of nooit
FEW (104)	62	29	13
FRG (40)	27	13	0
FBK (35)	23	9	3
FSW (30)	15	15	0
BMG (19)	13	6	0
FHKW (16)	9	5	2
FWB (6)	4	2	0
totaal (250)	153	79	18

Tabel 2. *Gebruikt u wetenschappelijke informatie (i.t.t. commerciële informatie)?*

Vraag 7 ging over Internet versus conventionele bronnen, waarbij de conventionele bronnen natuurlijk wel via Internet kunnen worden aangeboden. Ik denk dat dit onderscheid niet voor iedereen duidelijk is, de informatie komt immers uit hetzelfde kastje. Zoiets bleek ook uit de tekst van de vrije vragen. De vraag luidde: *Is informatie via Internet beter en/of beter te gebruiken dan informatie uit boeken en tijdschriften?*

De antwoorden waren: 15 % meestal, 66% soms en 19 % zelden of nooit.

Dit onderstreept de uitspraken dat het Internet nog bij lange na niet een vervanging van het traditionele informatieaanbod is.

	Meestal	soms	zelden of nooit
FEW (104)	24	66	14
FRG (40)	5	25	10
FBK (35)	5	21	9
FSW (30)	2	25	3
BMG (19)	1	13	5
FHKW (16)	0	12	4
FWB (6)	0	4	2
totaal (250)	37	166	47

Tabel 3. *Is informatie via Internet beter en/of beter te gebruiken dan informatie uit boeken en tijdschriften?*

Een goede zoekmethode is het gebruiken van bestaande lijstjes met bronnen. Vraag 8 was: *Maakt u gebruik van reeds verzamelde en gestructureerde Internetbronnen zoals DutchESS, verzamelingen van collega's en andere faculteiten/universiteiten?*

De antwoorden waren: 10% meestal, 40% soms en 50% zelden of nooit.

Duidelijk is dat deze bronnenlijstjes geen favoriete opties zijn. Wel werd vaak gevraagd naar zulke lijsten. Dit zou een goede dienstverlening van de UB zijn.

Deze verzoeken kwamen ook uit de FEW, de enige faculteit die op de faculteits-website een verwijzing heeft naar een flink aantal van dit soort bronnen. (Bij de andere faculteiten zijn er niet zulke lijsten aanwezig of goed versluierd zoals bij de FWB.)

	Meestal	soms	zelden of nooit
FEW (104)	10	42	52
FRG (40)	5	20	15
FBK (35)	4	12	19
FSW (30)	3	5	22
BMG (19)	3	9	7
FHKW (16)	2	6	8
FWB (6)	0	2	4
totaal (250)	27	96	127

Tabel 4: *Maakt u gebruik van reeds verzamelde en gestructureerde Internetbronnen zoals DutchESS, verzamelingen van collega's en andere faculteiten/universiteiten?*

Vraag 9 was *Bent u geabonneerd op discussielijsten voor uw vakgebied? Zo ja, hoe beoordeelt u daarvan de waarde?*

Deze vraag werd gesteld omdat het gebruiken van discussielijsten vaak wordt gepropageerd als onmisbaar bij het wetenschappelijk werk. De uitspraken waren duidelijk: 52 % n.v.t., 16% belangrijk en 32% niet zo belangrijk of waardeloos.

Discussiegroepen worden door meer dan de helft van de respondenten niet gebruikt en van de overigen vindt een derde het belangrijk.

Voor de mensen die er nog onbekend mee zijn zouden er lijsten met discussiegroepen moeten worden aangeboden, wellicht dat er iets bruikbaar voor hen bijzit.

Het zou ook nog mogelijk zijn voor deze laatste vragen een verdeling te maken naar succesvolle en niet-succesvolle zoekers. Voor wie dat relevant vindt zijn er programma's beschikbaar die de resultaten automatisch kunnen bewerken.

	n.v.t.	belangrijk	niet belangrijk
FEW (104)	56	17	31
FRG (40)	26	4	10
FBK (35)	18	7	10
FSW (30)	22	1	7
BMG (19)	6	1	12
FHKW (16)	0	7	9
FWB (6)	3	2	1
totaal (250)	131	39	80

Tabel 5. *Bent u geabonneerd op discussielijsten voor uw vakgebied? Zo ja, hoe beoordeelt u daarvan de waarde?*

3. De derde cluster vragen waren open vragen, die ongeveer door de helft van de respondenten zijn beantwoord.

Per faculteit en daarbinnen voor de succesvolle en voor de niet of minder succesvolle zoekers worden de antwoorden vermeld. Tussen haakjes het aantal keren dat hetzelfde bedoeld is. Tussen vierkantje haakjes mijn commentaar.

FEW, de succesvolle zoekers

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- grote zoekmachines (4)
- Altavista (5)
- Ilse.nl (2)
- Ask Jeeves (3)
- via sites van bekende organisaties (3)
- via startpagina (2) [welke? van de faculteit?]
- Metacrawler (3)
- Inomics (2)
- Resources for Economists RFE (2)
- via browsers [welke?]
- Google (2)
- Informs.org
- Blogs [??]
- Amnesty International news
- Prepublicaties vaktijdschriften [maar welke zoektechniek?]
- eigen bookmarks
- TU Delft [waarschijnlijk de startpagina]
- FTP-search [?]
- Virtuele boekwinkels (2)
- PICA, Picarta, OBN (3) [dit zijn geen zoektechnieken voor Internetgebruik, net als de volgende systemen; strikt genomen was dat ook niet gevraagd]
- Econlit (2)

- Web of Science (3)
- ERIC (2)
- WebEC
- NetEc
- Opmaat (2)
- Online Contents
- Ideal [?]
- Proquest (3) [waarschijnlijk hetzelfde als ProQuest]

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

De antwoorden zijn onderverdeeld in typische Internetproblemen, problemen met specifieke onderwerpen die misschien de betreffende vakreferent kan oplossen en problemen die op weer een ander terrein liggen.

Internetproblemen:

- meer helderheid in zoektechnieken [geannoteerde startpagina]
- één browser per discipline [doorzoekbare bronnenlijsten]
- bronnen van working papers [aanbieden van gateways op dit gebied]
- grijze literatuur op gebied van operations research / management science [idem]
- vertaling van Engelstalige zoekmachines [Is het zo erg? zie ook www.google.nl]
- methoden om het kaf van het koren te scheiden [Google; apart wetenschappelijk net]
- synoniemen en spellingherkenning
- verzameling links [aanbieden gateways]
- intelligent agents [voorlopig nog niet]

Specifieke problemen:

- World development indicators moeten gesaved kunnen worden in Excel format.

Andersoortige problemen en wensen:

- gemakkelijker toegang tot databanken [misschien wordt ook bedoeld zoeken in meerdere bronnen tegelijk?]
- tijdschriften (en tijdschriftartikelen) online (3)
- thuisgebruik bevorderen

FEW, de minder en niet-succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- Altavista (5)
- Ilse (5)
- Ask Jeeves (2)

- via websites van universiteit of organisatie (2)
- Search.nl en search.com
- Metacrawler
- Startpagina.nl
- NL-menu
- lijst van dHr Knoops [???
- via het NIWI
- Net-v.com / metazoeek/ [?]
- via discussiegroepen
- op z'n Jan Boerenfluitjes [!]
- Proquest (2)
- databanken UB
- Online Contents, helaas alleen titels
- ABI Inform

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

De antwoorden zijn onderverdeeld in typische Internetproblemen en problemen die op een ander terrein liggen.

Internetproblemen:

- meer voorlichting [pagina's met tips]
- duidelijke zoekmogelijkheden (3) [idem]
- betere afstemming tussen vraag en antwoord (3) [bijvoorbeeld google]
- bronnenlijsten
- onderscheid tussen wetenschappelijke en andere informatie (2) [wetenschappelijk Net]
- filter tegen hobbyisten [idem]
- zoekmachines geven teveel hits [betere zoektechnieken door pagina met tips]
- meer bronnen zoals Proquest voor wetenschappelijke artikelen

Andere problemen

- bibliotheekcatalogus toegankelijk maken met Z39.50 server (compabiliteit met Endnotes)
- lijst met alle tijdschriftartikelen (3)
- tijdschriften online

FRG de succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- bekende search engines (3)
- Alta vista
- Copernic
- Kluwer belastingsite

- via von Humboldt Univ. en via Aberdeen
- JOL
- Opmaat voor kamerstukken (2)
- PICA en Picarta (2)
- UB-databanken
- Telnet [?]

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

De antwoorden zijn onderverdeeld in typische Internetproblemen en problemen die op een ander terrein liggen.

Internetproblemen

- goede cursus [kan via startpagina]
- via UB naar bronnen zoeken [bronnen aanbieden]
- gericht zoeken naar wetenschappelijke onderwerpen [apart wetenschappelijk Net]
- zoektechnieken zijn onduidelijk [pagina met tips]
- handige zoekmachines beter bereikbaar [aanbieden via startpagina]

Andere problemen en wensen

- tijdschriften via Internet

FRG de minder en niet succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- bijhouden logboek
- surfen (2)
- JOL
- via bekende weg (3) [!?]
- eigen bookmarks
- Yahoo
- databanken UB
- NIWI, ook voor bestellen van artikelen en boeken

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

De antwoorden zijn onderverdeeld in typische Internetproblemen en problemen die op een ander terrein liggen.

Internetproblemen

- contextgevoelige hulp [iets voor concept maps]
- simpele zoekmethodes (2) [pagina met tips]
- bronnen zijn zoek [aanbieden van gateways]
- vakgericht zoeken [idem]
- index van discussielijsten [idem]

- index van tijdschriften op het web [idem]
- index van organisaties per vakgebied [idem]
- geen reclames [google, apart wetenschappelijk net]
- zoekmachines en indexen voor wetenschappelijke vakken [idem]

Andere problemen

- toegang tot cd's van zusterorganisaties [licentieprobleem]

FBK de succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- bekende zoekmachines (4)
- Altavista (5)
- Google
- bekende sites [welke?]
- Universiteitssite [welke?]
- Proquest (7)
- Endnote [?]
- Sherlock [is voor Mac]
- Pica (3)
- Web of Science (4)
- EconBase
- Informs.org (2)
- ABI/Inform (2)

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

De antwoorden zijn onderverdeeld in typische Internetproblemen en problemen die op een ander terrein liggen.

Internetproblemen

- zoektechnieken samenvatten [pagina met tips]
- effectieve technieken [idem]
- goede search engines aanbieden [idem]
- bronnen aanbieden [gateways aanbieden]

Andere problemen en wensen

- Proquest anders dan via ERNA aanbieden
- UB-catalogus en Proquest doorzoeken d.m.v. de connect optie in Endnote
- tijdschriften online (2)
- meer artikelen in databases
- kwantitatieve databases
- zoeken in meerdere databanken tegelijk

- verbeteren van de functie van de UB-site, zie: <http://wos.library.tudelft.nl/ciw.cgi>
[dit is Web of Science, die de UB/EUR ook aanbiedt!]

FBK de minder en niet succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- Alta vista (3)
- Ask Jeeves
- via uitgevers en bookstores
- Proquest
- ABI-Inform
- kranten [?]
- CBS

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

De antwoorden zijn onderverdeeld in typische Internetproblemen en problemen die op een ander terrein liggen.

Internetproblemen en wensen

- onderscheid tussen vakpublicaties en wetenschappelijke publicaties; zoekmachines leveren veel van het eerste en weinig van het tweede [subtiel onderscheid]
- effectievere technieken [pagina met tips]
- betere antwoorden bij de zoekvraag [concept maps]
- wetenschappelijke bronnen [aanbieden gateways]

Andere problemen en wensen

- tijdschriften online

FSW de succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- Altavista
- Yahoo
- Google
- bookmarks (2)
- vanaf bekende startpagina's
- Netscape search
- bekende organisaties zijn goed vindbaar
- Encycl Brit
- via bibliotheeksites van Nederlandse universiteiten
- CBS Statline

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

Internetproblemen

- tips voor effectief zoeken [pagina met tips]

FSW de minder en niet succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- Savvysearch.com
- via lijsten
- via bekende sites (voor overheidsinfo)
- Ask Jeeves
- browsen
- Picarta

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

De antwoorden zijn onderverdeeld in typische Internetproblemen en problemen die op een ander terrein liggen.

Internetproblemen

- simpeler zoeken (2) [aanbieden tips]
- effectiever zoeken via metacrawlers [metacrawlers op startpagina aanbieden]
- problemen met zoekmachine [aanbieden tips]
- weet niet wat beschikbaar is [gateways aanbieden]
- inventarisatie van complete boeken online [via Gutenbergproject?]
- bibliotheeksite zoals bij FSW [maar dan uitgebreider]

Andere problemen en wensen

- programma Sherlock voor PC (bestaat voor Mac)
- nieuwe aanwinsten van de UB zoeken is lastig omdat je geen jaar van uitgave aan kunt geven
- zoeken met één machine in gekoppelde bestanden [waarschijnlijk wordt bedoeld de UB-databanken]

BMG de succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- via NIWI
- bekende search engines (4)
- Altavista
- surfen vanaf bekende sites (4)
- Goliath

- Hotbot
- eigen bookmarks
- Pubmed
- Medline
- Web of Science
- Netscape search

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

Internetproblemen

- synoniemen en spelling
- working papers zijn moeilijk te vinden [aanbieden gateways]

BMG de minder en niet succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- Who [?]
- Nhi [National Health Institute?]
- NL-menu

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

- teveel treffers [aanbieden van tips]
- meer gericht kunnen zoeken [idem]
- kaf van het koren onderscheiden [google, gateways en apart wetenschappelijk net]
- zoeken met KI methodes [nog niet beschikbaar]

FHKW de succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- Hotbot
- Copernic 2000 (2)
- Yahoo
- Pica
- NCC (2)
- Online Contents
- Ask jeeves

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

De antwoorden zijn onderverdeeld in typische Internetproblemen en problemen die op een ander terrein liggen.

Internetproblemen

- lijst van sites van tijdschriften en dagbladen [aanbieden gateways]
- bibliografieën [idem]

Andere problemen en wensen

- e-mail-lijsten van EUR-medewerkers [bestaat al op de homepage van de EUR]

FHKW de minder en niet succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- Copernic
- Ask Jeeves
- Altavista
- Yahoo via homepage FHK

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

- bronnenlijstjes [aanbieden gateways]

FWB de succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- bekende zoekmachines
- Yahoo catalogus
- Boolean zoeken via zoekmachines
- Copernic 2000
- Meta search
- Gossip

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

- zoekmachines voor wetenschappelijk werk [google; apart wetenschappelijk Net]

FWB de minder en niet succesvolle zoekers.

Antwoorden op de vraag: *Heeft u reeds goede ervaringen met bepaalde zoektechnieken en welke zijn dat?*

- Alta vista
- Google
- surfen

De tweede vraag luidde: *Heeft u wensen op het gebied van zoektechnieken?*

- geselecteerde bronnen [aanbieden gateways]
- cursussen en tips [via startpagina].