

Een traject voor het selecteren van zoekmachinesoftware

Eric Sieverts, Monique Teubner

Universiteitsbibliotheek Utrecht – sector Innovatie & Ontwikkeling

Bij het woord "zoekmachine" denken veel mensen aan Google of Yahoo, de favoriete zoekhulpmiddelen voor het World Wide Web. Zoekmachines vormen echter ook het instrumentarium om bijvoorbeeld lokale documentcollecties of intranetten te kunnen doorzoeken. Software kiezen voor zo'n toepassing is niet altijd een triviale zaak. Anders dan bij gebruik van al op internet beschikbare zoekhulpmiddelen, zijn hier vaak grote investeringen mee gemoeid, zo al niet rechtstreeks in geld, dan toch in elk geval in te investeren tijd. Deze bijdrage beschrijft een best-practice voor dit proces dat grotendeels is gebaseerd op ervaringen met een dergelijk keuzetraject bij de Universiteitsbibliotheek Utrecht *. Dit keuzetraject zal dan ook als referentiekader worden gehanteerd.

De situatie bij de Universiteitsbibliotheek Utrecht

Bij de UB Utrecht was al geruime tijd een zoekmachine in gebruik voor het doorzoeken van metadata (titels, auteursnamen, trefwoorden, samenvattingen) van wetenschappelijke publicaties. Dit betreft publicaties - meest tijdschriftartikelen - waarvoor op basis van licenties of anderszins toegang is verkregen tot de full-text op de sites van diverse uitgevers, leveranciers of andere aanbieders. Op deze manier biedt het zoekstelsel via één uniform interface toegang tot de full-text artikelen bij al die aanbieders, alles bij elkaar intussen ruim 13 miljoen artikelen. Het Omega-stelsel waarbinnen deze zoekmachine functioneert, is meer dan alleen dit zoekstelsel. Gebruikers kunnen ook toegang krijgen tot al het elders toegankelijke tijdschriftmateriaal via een bladerfunctie door tijdschriftenlijsten en inhoudsopgaven. Verder omvat het een beheersstelsel waarin ook de administratie van de digitale collectie kan worden bijgehouden. De tot dusverre gebruikte zoekmachinesoftware werd door de leverancier helaas niet meer onderhouden of verder ontwikkeld. Omdat de software ook wat functionele tekortkomingen bleek te hebben, moest naar vervanging worden uitgezien.

* Een beschrijving van dit keuzeproces is te vinden in [Sieverts & Teubner 2006]

1. Inleiding

Er zijn diverse redenen waarom een eigen zoekmachine voor veel organisaties een belangrijke functie kan vervullen. De observatie dat veel organisaties beschikken over grote hoeveelheden aan (digitale) informatie die van levensbelang zijn voor het primaire proces en de kernactiviteiten van de organisatie, is bijna een open deur. Daarbovenop komt wat we het google-effect kunnen noemen. Mede dankzij Google is "zoeken" een alomtegenwoordige voorziening geworden. Er is daardoor een verwachtingspatroon ontstaan dat altijd en overal een zoekvenster aanwezig zal zijn ("the ubiquitous search box") en dat altijd en overal alles te vinden zal zijn ("ambient findability").

Dankzij Google worden daarbij ook enkele onuitgesproken randvoorwaarden aan die zoekmogelijkheden gesteld. Ze moeten een even simpel interface hebben als Google en je moet er even makkelijk relevante resultaten mee kunnen krijgen als met Google. Anders worden ze niet gebruikt. Het is echter maar de vraag of ook binnen een werkomgeving altijd eenvoudig aan zulke hooggespannen verwachtingen kan worden voldaan. Een intranet is immers heel wat anders dan het internet. Wat op internet werkt - bijvoorbeeld Google's pagerank mechanisme voor relevance ranking - hoeft dat op een intranet nog niet te doen. En ook eisen aan relevantie en volledigheid van zoekresultaten zijn in een werkomgeving meestal aanzienlijk anders dan in een consumenten-omgeving.

Dat keuze en implementatie van een zoekstelsel voor organisaties een ernstig te nemen onderneming is blijkt ook uit de resultaten van onderzoeken die regelmatig worden uitgevoerd door de Delphi Group, consultants voor IT-toepassingen in het bedrijfsleven. Uit een onderzoek van april 2006 bleek bijvoorbeeld dat meer dan 40% van de ondervraagde kenniswerkers meer dan 40% van hun zoektijd van meer dan 6 uur per week kwijt waren aan het doorploegen van irrelevante informatie [Reynolds 2006]. 52% was dan ook ontevreden met hun "search experience"; niet meer dan 3% was enthousiast over het zoeken op hun intranet. Voor elke organisatie is het dan ook een uitdaging te zorgen voor goede kwaliteit van het zoekinterface van de organisatie en meer algemeen voor de kwaliteit van de bijbehorende "user experience".

In dit artikel wordt daarom een beschrijving gegeven van een nauwgezet uitgevoerd keuzetraject voor het selecteren van zoeksoftware. Bij het lezen daarvan is een belangrijke beginvraag of een keuze altijd zo ingewikkeld is, dat een zo uitgebreid keuzetraject, met veel vaak tamelijk tijdrovende stappen, zoals hier beschreven, gevolgd moet worden. Er zijn immers heel uiteenlopende situaties denkbaar. De gewenste toepassing kan heel kleinschalig zijn, waarbij met een standaardoplossing via een goedkoop pakket kan worden volstaan, terwijl een andere toepassing juist zeer grootschalig is. Ook in dat geval kan soms met een standaard oplossing van een goed bekendstaand softwarepakket worden volstaan. Maar er zijn ook situaties waarin de wensen (of eisen) heel specifiek zijn voor materiaal en werkwijze bij de betreffende organisatie, zodat niet met een standaard oplossing kan worden volstaan.

Daarnaast kunnen er enorme verschillen zijn in initiële kosten. Er zijn pakketten waarvan de aanschaflicenties ver boven de €100.000 uitkomen, terwijl er ook gratis open-source zoekmachine-software is. Het verschil daartussen is dat in het eerste geval een kant-en-klare – maar nog wel naar specifieke toepassingseisen te tunen – oplossing geleverd kan worden, terwijl in het tweede geval de centrale zoekkern nog met allerlei andere modules geïntegreerd moet worden tot een omvattend systeem,

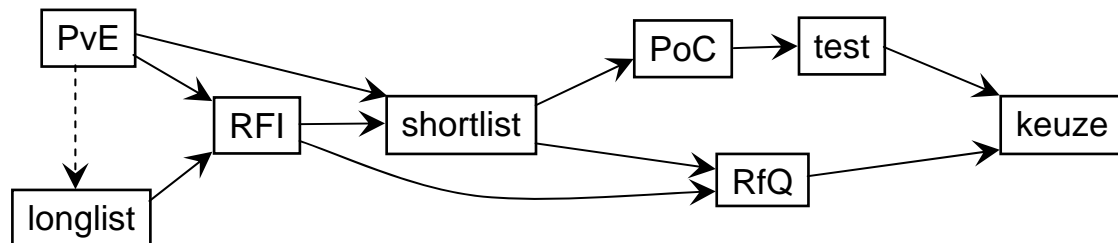
waarbij meestal nog veel programmeerwerk komt kijken. In tabel 1 worden wat voorbeelden genoemd van dergelijke sterk uiteenlopende soorten pakketten. De grote verschillen die tussen diverse situaties kunnen bestaan, betekenen dat het goed mogelijk is dat bepaalde stappen uit dit volledige traject in specifieke gevallen mogen worden overgeslagen of veel oppervlakkiger mogen worden uitgewerkt dan hier beschreven. Bij het lezen van deze tekst dient dit dus steeds in gedachten te worden gehouden.

Tabel 1: Voorbeelden van enkele zeer uiteenlopende soorten zoekmachine software, zoals die in 2006 op de markt zijn.

product	licentie	kenmerken	globale prijs-indicatie	URL
Google Desktop Yahoo! Desktop Ask Desktop Exalead Desktop	commercieel	kant-en-klaar; alleen voor single user; afgeleid van zoekmachine	0 0 0 0 - 10 ²	http://desktop.google.com/ http://desktop.yahoo.com/ http://about.ask.com/en/docs/desktop/ http://corporate.exalead.com/enterprise/
Aduna Autofocus	Open Source/ commercieel	kant-en-klaar + losse tools; alleen voor single user	0 - ?	http://www.aduna-software.net/
dtSearch	commercieel	van kant-en-klaar tot toolbox van single user tot web-based	10 ³	http://www.dtsearch.com/
Lucene	Open Source	alleen engine; applicatie daar omheen te bouwen	0	http://lucene.apache.org/
Google Mini	commercieel	kant-en-klaar; inclusief hardware	10 ³ - 10 ⁴	http://www.google.com/enterprise/
Google Search Appliance	commercieel	kant-en-klaar; inclusief hardware	10 ⁴ - 10 ⁵	http://www.google.com/enterprise/
Isys	commercieel	van kant-en-klaar tot toolbox	10 ⁴	http://www.isys.com.au/products/
Irion	commercieel	naar specificatie	10 ⁴ - 10 ⁵	http://www.irion.nl/dutch/products.html
Collexis	commercieel	naar specificatie	10 ⁴ - 10 ⁵	http://www.collexis.com/
Inxight	commercieel	geheel naar keuze	10 ⁴ - 10 ⁵	http://www.inxight.com/products/
Autonomy/Verity	commercieel	geheel naar keuze	10 ⁵ – up	http://www.autonomy.com/
Convera Retrievalware	commercieel	geheel naar keuze	10 ⁵	http://www.convera.com/
Fast	commercieel	geheel naar keuze	10 ⁵ – up	http://www.fastsearch.com/

2. Het totale keuzetraject

De stappen in een keuzetraject moeten voor een belangrijk deel opeenvolgend worden doorlopen. Om het hele proces niet onnodig te vertragen, kan op punten waar dat mogelijk is voor parallelle trajecten worden gekozen. Een typisch traject ziet eruit, zoals geïllustreerd in figuur 1.



Figuur 1 - Traject om tot keuze van software te komen

De stappen en tussenproducten in dit traject worden hier vast kort omschreven. In het vervolg van deze publicatie komen ze veel uitgebreider aan de orde.

- **PVE.** Op basis van de beoogde toepassing dient een liefst uitgebreid programma van eisen te worden opgesteld. Behalve een zorgvuldige analyse van het beoogde gebruik, speelt ervaring met tot dusverre gebruikte software en algemene kennis van de ontwikkelingen op retrieval-gebied, daarbij vaak een belangrijke rol.
- **Longlist.** Voor je verder kunt moet een marktverkenning zijn gedaan welke pakketten er op de markt zijn. Hoewel dit in principe parallel met het PVE zou kunnen, is het aan te raden dat PVE wel al in je achterhoofd te hebben om de longlist niet onnodig lang te laten worden.
- **RFI.** Van de pakketten uit de longlist wil je uiteraard allerlei gegevens weten en zeker hoe ze scoren op het PVE. Gebruikelijk is de leveranciers van de pakketten een *Request for Information* toe te sturen. Daarin wordt hen verzocht de daarin gestelde vragen te beantwoorden en aan te geven aan welke onderdelen van het PVE kan worden voldaan en voor welke onderdelen dat eventueel onder speciale condities mogelijk is.
- **Shortlist.** Op basis van de op het RFI ontvangen antwoorden, waarin zeker ook een eerste prijsindicatie moet worden gegeven, kan een selectie worden gemaakt van pakketten die het meest in aanmerking komen voor meer diepgaand onderzoek. Op deze lijst zullen er liefst niet meer dan vier of vijf overblijven.
- **POC.** Het is meestal aan te raden met de leveranciers van de overgebleven pakketten een *Proof of Concept* af te spreken. Op basis van het Programma van Eisen en met gebruik van een representatieve set gegevens uit het bestaande informatiesysteem, kan hen worden gevraagd zo goed mogelijk vergelijkbare

prototypes voor het zoekstelsel in te richten, zodat de overgebleven pakketten in een zo realistisch mogelijke situatie kunnen worden beoordeeld en vergeleken.

- **Test.** De door de leveranciers gerealiseerde prototypes dienen zorgvuldig beoordeeld te worden. Op basis van het PVE dienen daarvoor standaard testprotocollen te worden opgesteld, vooral om de functionele eigenschappen uit te testen. Anderzijds dient ook een zo goed mogelijke indruk te worden verkregen van de technische kant van de systemen.
- **RFQ.** Parallel aan het test-traject kan de leveranciers al een *Request for Quotation* – een verzoek om een offerte – worden gestuurd. Op basis hiervan kunnen vaak ook al initiële prijsonderhandelingen worden gevoerd.
- **Keuze.** Op basis van de resultaten van de *Proof of Concept*, van de uitonderhandelde prijs en vaak ook nog op grond van additionele overwegingen, kan tenslotte een definitieve keuze worden gemaakt.

3. Programma van eisen

3.1 Eisen aan zoekmachinesoftware

Onder 'Programma van Eisen' kunnen in de praktijk nogal uiteenlopende soorten documenten worden verstaan. Dat kan bijvoorbeeld samenhangen met de aard van de te selecteren software. Gaat het om een type toepassing waarvoor de software standaard al kant-en-klaar geconfigureerd moet zijn, of is het alleen maar bedoeld als een tool dat nog volledig naar eigen wensen ingericht moet kunnen worden? Gaat het om een implementatie van zo'n tool dat geheel geconfigureerd en voorgeprogrammeerd geleverd moet worden, of is het een tool dat lokaal voor allerlei uiteenlopende toepassingen geprogrammeerd moet gaan worden.

Zoekmachinesoftware zal in dit opzicht vaak een soort tussenpositie innemen. Er is wel concreet aan te geven welke zoekfunctionaliteit het tenminste moet bieden, maar alle details van de implementatie - bijvoorbeeld hoe de schermen er uit moeten zien - zullen vaak nog vrij gelaten worden. Een eis is dan natuurlijk wel dat de software voldoende flexibiliteit moet bieden deze details naar eigen wens te configureren en in te richten. Daarnaast zijn de technische randvoorwaarden waaraan de software moet voldoen natuurlijk wel op standaard wijze te specificeren. Zoals een belangrijk deel van die technische randvoorwaarden wordt bepaald door de lokale ICT-omgeving, zo zullen de functionele eisen direct voortkomen uit de aard van het materiaal waarvoor het systeem gebruikt moet worden en uit de manier waarop en de doelgroep waardoor het zoekstelsel gebruikt gaat worden.

De te doorzoeken content moet daarom meestal als beginpunt worden genomen. Dit kan eigen digitaal geproduceerd of gedigitaliseerd materiaal zijn, alsook van anderen ontvangen of aangeschaft digitaal materiaal. Dat materiaal kan allerlei vormen hebben: Word-files, PDF-documenten, webpagina's, gestructureerde XML-documenten, enzovoort. De aard van de content, naar inhoud en naar technische vorm, is meteen al een belangrijke bepalende factor voor hoe je er in wilt (en kunt) zoeken en dus welke functionele en technische eisen gesteld moeten worden.

3.2 Opstellen van een PVE

Voor het opstellen van een PVE hoeft meestal niet helemaal met een schone lei te worden begonnen. Soms kan al worden uitgegaan van een soortgelijk document bij een eerdere softwarekeuze. Soms kunnen lijsten met eisen van bevriende andere organisaties als uitgangspunt dienen. In andere gevallen zal de literatuur geraadpleegd moeten worden of kan een in de betreffende materie gespecialiseerd adviesbureau worden ingeschakeld. Daarnaast zullen er vaak aanvullende punten zijn die voortkomen uit problemen met tot dusverre gebruikte software. Eisen en criteria die voortkomen uit ervaringen met moderne zoeksystemen, vooral op internet, van de zoekspecialisten uit de organisatie spelen daarbij ook een belangrijke rol.

Eisen die voortkomen uit gebruikersonderzoek of algemeen onderzoek aan zoekgedrag, laten zich lang niet altijd vertalen in concrete eisen aan een zoekmachine. De meeste moderne zoeksystemen zijn namelijk zodanig flexibel te configureren dat, op basis van algemene functionaliteit, technische werking en mogelijkheden, voor de gebruikers nog allerlei interfaces en functionaliteiten gerealiseerd kunnen worden. Ervaring leert ook dat een bepaalde functionele eis vaak op technisch verschillende

manieren gerealiseerd kan worden. Functionele eisen moeten dus niet te sterk bij voorbaat al in termen van technische oplossingen worden geformuleerd.

Bij de formulering van een programma van eisen blijkt dat er - zowel op het terrein van functionaliteit, als bij de zuiver technische kant - vaak meer sprake is van wensen en beoordelingscriteria dan van alleen maar harde eisen. Dat illustreert dat bij elk punt in een PVE in principe nog de status en het belang ervan moeten worden aangegeven. Als alle punten uit een PVE werkelijk ‘harde’ eisen zouden zijn, zou waarschijnlijk nooit enig product overblijven dat aan een PVE voldoet. Er dient dus zorgvuldig en tamelijk selectief te worden bekeken wat werkelijk harde eisen zijn, waarop pakketten zonder meer afvallen als ze daaraan niet voldoen. Wat overblijft zijn dan eigenlijk “wensen” die in principe *nice to have* moeten zijn. Bij elk daarvan moet liefst een weegfactor worden vastgesteld die het belang van de betreffende wens aangeeft. Die weegfactoren zijn later van belang bij het maken van de selectie welke pakketten uit de oorspronkelijke lijst van mogelijke kandidaten in meer detail onderzocht moeten worden. Soms blijken er uiteindelijk ook nog punten over te blijven die alleen maar *nice to know* zijn en helemaal niet in de vorm van een eis of wens geformuleerd kunnen worden, zoals bijvoorbeeld gegevens hoe vaak nieuwe releases van een pakket uitkomen.

Als voorbeeld volgt hieronder een uitwerking van de rubrieken zoals die in een PVE voor zoekmachinesoftware kunnen voorkomen, met daarbij enige toelichting. Daarbij komen functionele eisen, technische eisen en leverancierscriteria aan de orde. De meeste aandacht zal daarbij uitgaan naar de functionele eisen omdat die heel specifiek zijn voor het type software waar het in dit artikel om gaat. In tabel 2 wordt daarom ook een mogelijke verdere uitwerking van de functionele eisen gegeven, die als checklist voor een werkelijk programma van eisen kan dienen.

3.3 Functionele eisen - zoeken

Welke zoekmogelijkheden een systeem kan bieden wordt voor een belangrijk deel bepaald door de methoden die de zoekmachine kan toepassen om de te doorzoeken informatie te indexeren. Toch zal de gewenste zoekfunctionaliteit steeds het uitgangspunt moeten zijn bij het formuleren van een programma van eisen. Enerzijds zullen die vaak bestaan uit klassieke eisen voor Booleaans zoeken en daaraan gerelateerde tamelijk deterministische zoekfuncties. Anderzijds creëren internet-zoekmachines steeds meer het verwachtingspatroon dat systemen ook meer probabilistische zoektechnieken kunnen toepassen, waarbij zoekvragen uit een rijtje woorden mogen bestaan en zoekresultaten worden geordend op basis van hun berekende (kans op) relevantie. Daarbij bestaat dan ook vaak de wens het gewicht en het onderlinge belang van de verschillende factoren die daarbij meespelen, te kunnen aanpassen. Het blijkt trouwens vaak erg lastig om zelfs voor het eigen materiaal objectieve criteria vast te stellen voor de beoordeling van de relevantie-ordening van een zoekmachine.

Bij deze zoektechnieken hoort ook vaak de wens dat een al gevonden relevant document als uitgangspunt voor een “more like this” zoekactie genomen kan worden. Daarnaast kan ook de toepassing van algemene taaltechnologische technieken gewenst zijn. Denk daarbij aan fuzzy search, aan reductie van woorden tot hun morfologische woordstam en aan het automatisch uitbreiden van zoekvragen met via een semantisch netwerk of een thesaurus gevonden verwante begrippen.

3.4 Functionele eisen - feedback

Voor het verbeteren van zoekresultaten bestaan diverse feedback mechanismen. Bij informatie die ook meer geformaliseerde metadata bevat, kan zogenaamd "parametrisch zoeken" een handige methode zijn om zoekresultaten in te perken. Bij het resultaat van elke zoekactie wordt dan meteen al getoond hoeveel van de gevonden informatie uit verschillende jaren afkomstig is, hoeveel daarvan tot bepaalde documentsoorten behoort, hoeveel door een bepaalde auteur is geschreven, enzovoort. Zo weet de gebruiker tevoren op welke formele gegevens de zoekactie zinvol ingeperkt kan worden.

Voor minder geformaliseerde gegevens kunnen suggesties voor aanvullende zoekwoorden uit de gevonden documenten worden gegenereerd. Daarbij worden meestal statistische technieken gebruikt om de meest kenmerkende termen uit de tekst van die documenten te halen. Dit is een meer vraaggerelateerd alternatief voor het automatisch toevoegen van woorden uit woordsystemen als een semantisch netwerk of een thesaurus.

Wel moet de afweging worden gemaakt welke van al deze mogelijke zoektechnieken ook werkelijk zullen worden geïmplementeerd en aan de gebruikers aangeboden. De wensen en verwachtingen van de informatie-professionals ten aanzien van een zoekstelsel hoeven niet altijd dezelfde te zijn als die van de (eind)gebruikers in een organisatie.

Bij de presentatie van zoekresultaten is in de eerste plaats terugkoppeling naar de gebruiker van belang. Voor de hele resultaatset gaat het om terugkoppeling waarop eigenlijk was gezocht, en bij elk afzonderlijk item om terugkoppeling waarom juist dat document is gevonden. Bij probabilistisch zoeken betekent dat een aanduiding welke van de gevraagde zoektermen in het document aanwezig zijn; bij elk soort zoeken bovendien een aanduiding waar ze in het document staan, middels highlighting of een KWIC-presentatie (Keyword in context).

Daarnaast zijn er vaak eisen hoe zoekresultaten geordend moeten kunnen worden. Daarbij is het voor systemen die aan relevantie-ordening doen, niet zonder meer duidelijk hoe ook op een ander element (bijv. datum) geordend moet kunnen worden. Het is immers slecht te verkopen dat een document dat weliswaar heel recent is, maar een heel lage relevantiescore heeft, dan toch bovenaan zou komen te staan.

Bij het bekijken van zoekresultaten kan het verder handig zijn relevante documenten tijdelijk in een "winkelwagentje" te kunnen plaatsen, om aan het eind van een zoeksessie alle zo verzamelde resultaten in één keer te kunnen downloaden of uitprinten.

3.5 Functionele eisen - gebruikersinterface; personalisatie

Wat betreft het gebruikersinterface is een eis van "gebruikersvriendelijkheid" altijd moeilijk objectief te operationaliseren. Dankzij de vele webtoepassingen wordt daar, als onderdeel van het algemenere begrip "usability", veel over geschreven. Hoewel er daarbij nadruk op wordt gelegd dat de gebruiker en diens zoektaak daarbij als uitgangspunt moeten worden genomen, is er toch ook wel enige consensus over de eisen die op dat terrein gesteld moeten worden [Resnick 2006, Quesenbery 2003, Rappaport 2005]. Als zulke eisen in een PVE worden opgenomen, zullen ze echter alleen betrekking kunnen hebben op standaard meegeleverde interfaces. In veel gevallen is het juist veeleer van belang gebruikersinterfaces geheel naar eigen wensen te kunnen aanpassen. Daar kunnen diverse redenen voor zijn:

- om het te kunnen aanpassen aan specifieke doelgroepen,
- om het te kunnen aanpassen aan een specifiek type content,
- om te kunnen aansluiten bij bepaalde uitgangspunten van de organisatie (huisstijl, eenvoud, ...),
- om geen problemen te ondervinden in geval van "voortschrijdend inzicht" op enig van deze terreinen.

Pas bij zelf bouwen van interfaces komen die usability-eisen dan weer aan de orde.

Bij veel organisaties zullen ook personalisatie-functies van belang zijn. Daarbij kan gedacht worden aan standaard instellingen voor het zoekproces zelf, zoals inperkingen op een bepaald deel van de collectie, maar ook aan mogelijkheden allerlei zaken meer permanent op te slaan. Naast een tijdelijk "winkelwagentje", kunnen ook enkele permanente "boekenplanken" van nut zijn voor gevonden materiaal dat vaker benodigd is. Evenals een collectie "favoriete" zoekvragen waarin succesvolle zoekvragen kunnen worden bewaard, om later opnieuw te laten uitvoeren om te zien of nieuwe relevante informatie aan het systeem is toegevoegd. In dat laatste geval kan dan ook meteen gedacht worden aan automatische attendering met een door de gebruiker in te stellen frequentie.

3.6 Functionele eisen - indexering

Zoekmogelijkheden worden voor een belangrijk deel bepaald door de wijze waarop de te doorzoeken informatie door de zoekmachine geïndexeerd wordt. Maar bepaalde zoekfuncties kunnen soms ook op meer manieren technisch gerealiseerd worden. Daarom hoort een programma van eisen niet onnodig te worden dichtgetimmerd met teveel (niet strikt noodzakelijke) eisen aan de indexering. Hooguit kan een eis dat het systeem relevance ranking moet toepassen, onder indexering terugkomen met de algemene eis dat een indexeermethode moet worden toegepast die dit mogelijk maakt. Wel horen onder het kopje "indexering" uiteraard de volgende soorten eisen;

- welke documenttypes standaard geïndexeerd moeten kunnen worden, waarbij misschien ook tekstgegevens uit (relationele) databasesystemen,
- dat de wijze van indexering – eventueel per veld – geconfigureerd en gespecificeerd kan worden,
- dat wellicht bepaalde leestekens meegeïndexeerd moeten worden,
- dat in het te indexerende materiaal diacrieten in verschillende codering herkend moeten worden.

Verder vallen hieronder eisen die te maken hebben met "spider" functionaliteit, de mogelijkheid om ook de inhoud van geselecteerde websites – zowel HTML-pagina's als PDF- of andere documenten – te indexerende en full-text doorzoekbaar te maken.

Tabel 2: Beperkte checklist voor functionele eisen voor zoekmachinesoftware *

Zoeken	
Booleaanse zoekmogelijkheid	zoekwoorden met operatoren AND, OR en NOT kunnen combineren, eventueel met gebruik van geneste haakjes voor meer complexe zoekvragen
sets combineren	eerder gemaakte resultaatsets achteraf weer met Booleaanse operatoren kunnen combineren
zoeken binnen zoekresultaat	binnen een al verkregen zoekresultaat een vervolg-zoekactie doen (impliciet AND daarmee)
nabijheidszoeken	mogelijkheid aan te geven dat gevraagde woorden binnen een gespecificeerde afstand van elkaar en (eventueel) in gegeven volgorde moeten voorkomen
exacte zin	mogelijkheid aan te geven dat de gevraagde woorden exact zo als zin moeten voorkomen
truncatie	gebruik van "wildcards" aan einde van zoekwoord of voor intern maskeren van variabele letters
fuzzy zoeken	zoeken naar woorden die in spelling of uitspraak sterk lijken op het zoekwoord
word stemming	voor het zoeken worden woorden tot hun morfologische woordstam gereduceerd (regels hiervoor voor alle relevante talen beschikbaar)
vraagexpansie met synoniemen	automatische expansie van de zoekvraag met synoniemen of anderszins inhoudelijk gerelateerde woorden uit semantisch netwerk, synoniemenlijst, thesaurus, e.d.
speciale zoekfuncties ook UIT te zetten	word stemming, fuzzy zoeken, vraagexpansie e.d. zijn niet in elke situatie gewenst
zoeken in specifieke velden	bij enigszins gestructureerde gegevens mogelijkheid specifiek binnen bepaalde (metadata-)velden te zoeken
limitering op geformaliseerde gegevens	mogelijkheid zoekactie in te perken op standaardwaarden uit voldoende groot aantal metadata-velden (zie ook "parametric search")
probabilistische of best match zoektechniek	zoektechniek waarbij de zoekresultaten op hun berekende waarschijnlijke mate van relevantie worden geordend, waarbij de resultaten vaak ook niet volledig aan de zoekvraag hoeven te voldoen
relevantiefactoren in te stellen	mogelijkheid om het (onderling) belang van de voor relevantieordening gebruikte factoren in te stellen
instelbare relevantiedrempel	instelbaar bij welk relevantiepercentage de resultatenlijst moet worden afgekapt
probabilistisch en Booleaans zoeken te combineren	op dezelfde collectie of ook binnen één zoekvraag kunnen probabilistisch en Booleaans zoeken gecombineerd worden
"query by example"	al gevonden (relevant) document als uitgangspunt laten dienen om daarop gelijkende documenten te vinden
zoekgeschiedenis	overzicht van tijdens een sessie gestelde zoekvragen, eventueel met vermelding van gevonden aantallen

Interactieve zoekvraagverbetering	
parametric search	zoekresultaat is al uitgesplitst naar inhoud van bepaalde metadata velden; bijv. publicatiejaren, vakgebieden, documentsoorten, auteursnamen, e.d.
zoektermsuggesties	op basis van bijv. statistiek worden mogelijk zinvolle termen uit het zoekresultaat afgeleid, voor expansie of inperking van de zoekvraag
mogelijkheid relevante resultaten te markeren	aanvullende termen worden alleen aan gemarkeerde resultaten ontleend; gewicht van in gemarkeerde resultaten voorkomende termen wordt verhoogd
Presentatie van zoekresultaten	
aanpasbaar presentatieformat voor korte en uitgebreide presentatie	te specificeren op welke wijze welke velden en welke zoekvraaggerelateerde gegevens getoond moeten worden
volledige resultatenlijst te bekijken	ook bij lange resultatenlijsten alle gevonden resultaten te tonen
KWIC presentatie	presentatie van die delen van gevonden documenten waarin de zoekwoorden voorkomen
zoekresultaat te sorteren	zoekresultaten te sorteren op relevante criteria (relevantie, publicatiejaar, auteursnaam, ...)
"winkelwagen"-functie	mogelijkheid tijdens een zoeksessie individuele documenten te bewaren, om die aan het eind van de sessie te downloaden, printen of mailen
Gebruikersinterface	
gebruikersvriendelijk standaard-interface	standaard meegeleverd interface voldoet aan huidige opvattingen over usability van zoeksystemen
geheel aanpasbaar gebruikersinterface	interface is geheel aan specifieke wensen aan te passen, bijv. via tussenlaag tussen webgebaseerde schermen en interne software-engine
Personalisatie	
persoonlijke gebruikersprofielen	profielen waarin gegevens over attenderingen, persoonlijke voorkeuren, virtuele boekenplanken, e.d.
standaard instellingen	te bewaren standaardinstellingen voor zoekvoorkeuren, te doorzoeken (deel-)collecties e.d.
persoonlijke boekenplank	mogelijkheid gevonden documenten te bewaren op een persoonlijke, al dan niet in categorieën in te delen virtuele boekenplank
bewaren van zoekvragen	zoekvragen opslaan om ze later nogmaals te kunnen laten uitvoeren
automatische attendering	attendering (per e-mail of ander medium) op nieuwe informatie die voldoet aan opgeslagen zoekvragen / zoekprofielen
attenderingsfrequentie instelbaar	in te stellen hoe vaak per dag, week of maand men geattendeerd wil worden
meer mail-adressen voor attendering op te geven	meer mailadressen om groepsattendering mogelijk te maken

Indexering	
herkenning van belangrijke file-formats	herkent tekst in ASCII, Word, PDF, XML, HTML en andere vereiste file-formats; eventueel ook in de inhoud van velden in (relationele) database-systemen
indexering van losse woorden en samengestelde begrippen	samengestelde termen moeten herkend kunnen worden en zowel als geheel, als op losse woorden geïndexeerd kunnen worden
indexering van bijzondere tekens	leestekens (& ' / -) desgewenst te indexeren; diacrieten in verschillende coderingswijze herkenbaar; equivalentie met ongeaccentueerde letters
configureerbare indexeerregels	(eventueel per veld) instelbaar o.a.: indexering van samengestelde begrippen (zoals in trefwoorden of auteursnamen), toepassing van word-stemming, indexering bijzondere tekens e.d.
veldherkenning	om veld-specifiek zoeken mogelijk te maken
aanpasbare stopwoordenlijst	lijst van niet te indexeren woorden, liefst afzonderlijke lijsten voor verschillende talen
indexering t.b.v. best-match zoeken	moet indexeringsmethode toepassen die best-match of probabilistisch zoeken mogelijk maakt
indexering t.b.v. fuzzy zoeken	moet indexeringsmethode toepassen die fuzzy zoeken mogelijk maakt
incrementele update van indexen	bij toevoegen van nieuwe gegevens geen gehele nieuwe index bouwen
goede indexeer performance	zowel wat betreft snelheid als geheugenbeslag moet aan bepaalde minimum eisen voldaan worden
aanwezigheid internet-spider	via http-protocol op internet bereikbare webpagina's en documenten indexeerbaar
start-URL's specificeerbaar	op te geven lijst van URL's waar spider zijn indexering moet beginnen
indexeerdiepte instelbaar	te specificeren hoe diep in opgegeven sites geïndexeerd moet worden en/of hoeveel opvolgende links buiten die sites gevolgd moeten worden
update frequentie	te specificeren hoe frequent index voor gespiderde bronnen te updaten
metadata-records bij web-documenten	mogelijkheid web-documenten te linken met aparte metadata-records

- * Deze functionele eisen hebben betrekking op de functionaliteit ten behoeve van de (eind)gebruikers van het systeem. Eisen voor de aan beheerders van het systeem te bieden functionaliteit, zijn hier gerangschikt onder de technische eisen

3.7 Technische eisen en leverancier

Algemene technische eisen aan zoekmachinesoftware wijken voor het grootste deel weinig af van de eisen voor veel andere soorten software. Hier zullen we dan ook niet veel verder gaan dan een summiere aanduiding van de belangrijkste categorieën eisen.

Algemene eisen hebben onder meer betrekking op:

- operating system en hardware
- documentatie en ondersteuning

Een aantal categorieën technische eisen en criteria is direct aan de software zelf gerelateerd:

- gemak van out-of-the-box oplossing
- GUI voor systeem-configuratie
- gemak van configureerbaarheid van speciale wensen
- gebruik van algemeen aanvaarde standaarden en protocollen
- toegankelijkheid voor en connectivity met andere softwaresystemen
- mogelijkheden voor migratie naar een ander systeem
- beschikbaarheid van ontwikkel-tools
- solide regeling van toegang & beveiliging
- mogelijkheden tot "crash recovery"
- performance ten aanzien van indexering
- performance ten aanzien van responsetijden bij meer gelijktijdige gebruikers en complexe zoekvragen
- scalability
- gemak van aanpasbaarheid van het gebruikersinterface

Andere soorten eisen hebben betrekking op hardware, zoals:

- analyse & monitoring van workload, CPU-gebruik, geheugengebruik, transacties, e.d.
- application tuning voor workload, schijfverkeer, e.d.
- application modelling & sizing
- demand management

Ook aan de leverancier wordt meestal een aantal eisen gesteld. Die eisen en achterliggende criteria horen onder meer tot de volgende categorieën:

- betrouwbaarheid van het bedrijf en te verwachten continuïteit van het product en de dienstverlening
- te verlenen ondersteuning en beschikbare documentatie
- verzorgen van gebruikerstraining (voor de technisch verantwoordelijken)
- aard van de softwarelicentie
- rechten bij failliet gaan van de leverancier, bijvoorbeeld via een escrow-regeling
- gehanteerde tariefstructuur
- regeling van onderhoud en nieuwe versies
- welke toonaangevende andere gebruikers er zijn

Voor een echt PVE moeten – afhankelijk van de precieze situatie – de meeste van deze algemene punten nog verder worden uitgewerkt in meer specifieke eisen en criteria.

4. Van longlist naar shortlist

4.1 De longlist

Bij het begin van een keuzetraject is het handig als binnen de organisatie al enige kennis aanwezig is van de belangrijkste producten die op de markt zijn. Dat kan al een eerste aanzet geven voor een uitgebreide lijst met mogelijke kandidaten, voordat echt verder marktonderzoek gedaan gaat worden. Anders zijn op internet wel overzichten met beschrijvingen te vinden, bijvoorbeeld op de website van "Searchtools". Ook bij grote adviesorganisaties als Gartner, Forrester of Delphi-group is dergelijke informatie te vinden, vaak al voorzien van kwaliteitsoordelen. Hoewel hun volledige rapporten meestal peperduur zijn, zijn op hun websites vaak wel voldoende nuttige samenvattingen daarvan te vinden. Een andere aanpak is om een aantal experts naar hun persoonlijke favorieten te vragen, al zal men niet altijd bereid zijn dergelijke informatie gratis te verschaffen. Ook kan een marktverkenning helemaal uitbesteed worden aan een gespecialiseerd adviesbureau. In de praktijk wordt vaak een combinatie van dergelijke aanpakken toegepast. Zo hebben we voor het keuzetraject bij de Universiteitsbibliotheek Utrecht de zelf opgestelde longlist aangevuld met enkele leveranciers op basis van de expertise van het door ons ingehuurde adviesbureau Search Expertise Centrum. Op die lijst stonden uiteindelijk dertien producten.

4.2 Request for Information

Een volgende stap is om er zo goed mogelijk achter te komen welke producten uit de longlist aan je PVE voldoen – en liever nog welke dat het beste doen. De oplossing om dat allemaal zelf in de praktijk te testen is meestal onmogelijk. Ook is het tamelijk onwaarschijnlijk dat iemand anders dat toevallig al voor een soortgelijk PVE heeft uitgezocht, en dat die gegevens ook nog publiekelijk beschikbaar zijn. Daarom wordt er meestal voor gekozen om aan alle leveranciers op de longlist een Request for Information te sturen, waarin onder meer wordt gevraagd antwoord te geven op alle punten uit het PVE.

Wanneer een adviesbureau wordt ingehuurd, beschikt dat soms al over een database met de belangrijkste karakteristieken van de meest bekende producten. Niettemin zal dat meestal nog aangevuld moeten worden met specifieke gedetailleerde criteria uit het eigen PVE, alsmede met gegevens van leveranciers die nog in die database ontbreken. Ook zal het meestal nodig blijken de wel al beschikbare gegevens te updaten voor de meest recente versies van de betreffende producten. Een adviesbureau zal dus meestal zelf zo'n RFI voor een klant rondsturen.

Zo verzamelde gegevens dienen vervolgens nog kritisch tegen het licht te worden gehouden, omdat veel leveranciers geneigd zijn bijna elke vraag met JA te beantwoorden. Vergelijking met antwoorden op andere vragen of met technische specificaties uit standaard meegestuurde documentatie, kan dergelijke antwoorden vaak enigszins relativeren. Ook moet soms expliciet worden nagevraagd of het antwoord wel precies betrekking heeft op datgene wat in het PVE werd bedoeld. Op basis van de response op het RFI zullen vaak zonder meer al enige producten afvallen. Daar kunnen uiteenlopende redenen voor zijn. Een leverancier wil bijvoorbeeld niet de moeite nemen alle criteria uit het PVE te beantwoorden, omdat hij bij voorbaat inschat dat het beschikbare budget niet voldoende is om het betreffende product aan te schaffen. Een ander geeft op herhaalde verzoeken zelfs

helemaal geen reactie. Weer een ander laat weten zijn product (nog) niet in Europa op de markt te brengen. Nog weer een ander product blijkt wellicht door een concurrent overgenomen die alleen met zijn eigen oorspronkelijke product wil meedingen. Een bij het RFI verkregen prijsindicatie die (ver) boven het beschikbare budget uitkomt, moet nog niet te snel aanleiding zijn het product al definitief af te laten vallen. Veel leveranciers werken namelijk niet met heel precieze tarieflijsten, en laten vrij veel ruimte voor onderhandeling of laten de uiteindelijke prijs afhangen van allerlei technische details die in dit stadium meestal nog niet zijn vastgelegd.

Op basis van de essentiële eisen uit het PVE (*de "must have"*) en de weegfactoren bij de andere eisen (*de "nice to have"*) kan vervolgens zo goed mogelijk gekwantificeerd worden hoe de verschillende producten scoren.

4.3 Beoordeling

De uiteindelijke selectie welke producten uit de longlist overblijven, kan gebaseerd zijn op een volstrekt kwantitatieve berekening, bijvoorbeeld via een spreadsheet, waarmee je de afweging van de producten "meetbaar" probeert te maken. Wel geldt hierbij de waarschuwing je niet te verliezen in pseudo-kwantificeerbaarheid. Daarom kunnen de wegingen van de verschillende eisen of wensen soms ook wel wat lossier als kwalitatieve indicaties worden gehanteerd. In het bijzonder zullen "meer functionaliteit" en "minder kosten" in dit stadium waarschijnlijk nog niet echt objectief – laat staan kwantitatief – tegen elkaar afgewogen kunnen worden. Vaak zal er dan ook voor worden gekozen om in de resulterende shortlist zowel producten op te nemen die functioneel het best lijken te presteren, als producten met een gunstige prijs/prestatie verhouding, enigszins zoals de "voordelige keus" van de Consumentenbond.

In gevallen waarin beslist een low-budget oplossing wordt nagestreefd, zal het meestal niet mogelijk zijn leveranciers een RFI te laten beantwoorden. De te verwachten lage winstmarges zullen hen niet toestaan daar tijd aan te besteden. In die gevallen zal men zelf zo betrouwbaar mogelijke gegevens moeten verzamelen uit productspecificaties, documentatie, besprekingen of tests in de literatuur en ervaringen van andere gebruikers. Ook is het voor dat type software meestal mogelijk gratis demoversies te downloaden en die zelf in de praktijk uit te proberen. Dat komt dan meteen in de plaats van de in de volgende paragraaf beschreven Proof of Concept die vaak voor duurdere producten wordt georganiseerd.

5. Proof of Concept

5.1 Verdere analyse

De ervaringen leert dat - ook correcte – respons op de vragen uit een RFI, en daarmee op de eisen uit het PVE, nog niet altijd garandeert dat een zoekstelsel echt precies kan wat de opdrachtgever wil. In de oorspronkelijk in Utrecht gebruikte zoeksoftware bleek bijvoorbeeld dat de normaliter uitstekende probabilistische *relevance ranking* volledig onbruikbaar werd, zodra ook truncatie van zoektermen of fuzzy zoeken werden toegepast. Dat illustreert dat een stelsel dat aan alle eisen afzonderlijk uitstekend voldoet, bij een bepaalde combinatie daarvan, soms toch nog slecht kan presteren. Daarom zijn nog een paar vervolgstappen nodig om binnen de shortlist tot een gefundeerde verdere keuze te komen.

- Een eerste stap daartoe is de antwoorden op de RFI nog kritischer te analyseren dan ten behoeve van de eerste selectieronde al was gebeurd. Zo nodig dient bij de leveranciers dan nog preciezer navraag te worden gedaan over onduidelijke details of tegenstrijdig lijkende uitspraken.
- Een tweede stap is demonstraties te laten geven door de leveranciers van de overgebleven pakketten. Dat dient te gebeuren op basis van strikte voorwaarden voor wat beslist getoond moet worden. Eisen daarvoor zullen gewoonlijk voortkomen uit het PVE en uit de nog overgebleven onzekerheden uit het RFI. Uniformiteit in de aan de leveranciers te stellen eisen is van belang om de gedemonstreerde pakketten makkelijker onderling te kunnen vergelijken en beoordelen.
- Een derde stap zal vaak nog zijn andere gebruikers van de betreffende pakketten te bezoeken en kritisch naar hun ervaringen te vragen. Bij die gebruikers moeten de pakketten dan wel op een wijze worden toegepast, die redelijk representatief is voor de beoogde eigen situatie. Ook hier zullen vaak onzekerheden met betrekking tot het RFI aan de bevraging ten grondslag liggen. Tevens kan men zo meer te weten komen over de ondersteuning bij implementatie en de kwaliteit van de helpdesk.

5.2 Proof of Concept

Omdat uit deze drie stappen vaak nog geen alomvattende beeld van de pakketten naar voren komt, en ook om zeker te weten hoe de systemen met het eigen materiaal presteren, verdient het aanbeveling de overgebleven leveranciers ook een *Proof of Concept* te laten verzorgen. Dit houdt in dat de leveranciers wordt gevraagd prototypes te bouwen op basis van de belangrijkste eisen uit het PVE, met daarin een voldoende groot en representatief gedeelte van het eigen materiaal. Dit is zeker aan te raden als het PVE specifieke eigen wensen bevat, zodat geen standaard-implementatie mogelijk lijkt, en als het om tamelijk grote hoeveelheden gegevens gaat. Een goed opgezette Proof of Concept (POC) kan dan een betrouwbare indruk geven hoe goed de gewenste applicatie met de betreffende software te realiseren zal zijn.

In aanvulling op de zojuist gemaakte opmerkingen over te implementeren functionaliteit en doorzoekbaar te maken materiaal, moet worden bedacht dat bij het organiseren van een POC ook eisen van uniformiteit spelen, om de resultaten voldoende vergelijkbaar te maken:

- Hoewel alle pakketten hun eigen specifieke kenmerken en mogelijkheden hebben, moet er naar gestreefd worden de prototypes zo veel mogelijk vergelijkbaar te laten zijn wat betreft geïmplementeerde functionaliteit.

- In alle prototypes dient dezelfde content doorzoekbaar gemaakt te zijn.

Daarnaast dient uiteraard ruimte te worden gelaten om de individuele sterke punten van elk van de producten ook aan de orde te laten komen in de POC. Omdat juist in wat complexere of grootschaliger situaties gebruikersinterfaces grotendeels zelf ontworpen zullen worden, zal een beoordeling van het interface vaak geen onderdeel uitmaken van de POC. Wel zal de beoordelaars van de systemen een minimale mate aan gebruiksvriendelijkheid geboden moeten worden. Die testers moeten dan bovendien goed weten dat ze niet geacht worden zich te laten invloeden door verschillen tussen de interfaces van de bekeken prototypes. Ook moet het dan zeker zijn dat het gebruikersinterface voldoende flexibel aan alle mogelijke eisen is aan te passen.

Bij grootschalige of complexe POCs speelt wel de vraag of leveranciers die zo maar gratis willen verzorgen. Grote bedrijven zullen hiertoe meestal wel bereid zijn. Voor kleinere bedrijven zal het inrichten daarvan, zonder gegarandeerd uitzicht op latere inkomsten, een minder vanzelfsprekende investering zijn. Onderhandelingen over voorwaarden en eventuele kosten horen dan ook tot de voorbereiding.

Proof of Concept bij de Universiteitsbibliotheek Utrecht

Op basis van de resultaten uit het Request for Information en advies van ingehuurde specialisten, was besloten verder te gaan met drie leveranciers: de bekende marktleiders *Verity* en *Autonomy* (op dat moment nog concurrenten) en de Nederlandse producent *Irion*. Voor beide eerstgenoemde was dat op basis van hun totale score op het hele PVE, voor Irion op basis van vooral het functionele deel van de eisen, in combinatie met te verwachten lagere kosten. Hen is gevraagd een Proof of Concept te verzorgen, om te bewijzen dat ze aan de belangrijkste functionele en technische eisen konden voldoen en dat hun zoekresultaten bevredigend zouden zijn voor grote hoeveelheden materiaal zoals in het Omegasysteem. Daartoe bevatte elk van de drie zoeksystemen (dezelfde) 10% van de ruim 10 miljoen artikelen waaruit de totale Omega-collectie op dat moment bestond. Die artikelen waren representatief voor de totale collectie, zowel ten aanzien van de uitgevers van de artikelen als hun vakgebied en jaar van uitgave. Ter vergelijking waren dezelfde 1 miljoen documenten ook doorzoekbaar gemaakt met de tot dusverre gebruikte oude zoekmachine.

De systemen moesten zo worden geïmplementeerd dat ze dezelfde basisfunctionaliteit boden, waarin de belangrijkste functionele eisen uit het PVE ten aanzien van het zoeken waren verwerkt. Er moest dus zowel geavanceerd probabilistisch als klassiek booleaans gezocht kunnen worden. Verder moesten onder meer truncatie, word-stemming, fuzzy-zoeken, veld-specifiek zoeken en inperkingen op jaren en vakgebieden mogelijk zijn. Het zoekinterface hoefde nog niet aan specifieke eisen voor gebruiksvriendelijkheid te voldoen, aangezien later toch een nieuw interface gebouwd zou worden.

Er is geen moeite gedaan de drie systemen op volledig vergelijkbare hardware-systemen te laten inrichten. Deze test was immers niet bedoeld als benchmark voor de performance. Bovendien kan bij de meeste systemen achteraf nog zoveel worden aangepast dat resultaten van deze prototypes toch onvoldoende representatief zouden zijn voor de performance van de uiteindelijke systemen. De drie leveranciers hebben daarom hun systemen elk op eigen hardware geïmplementeerd en via internet toegankelijk gemaakt, zodat onze testers er op de eigen werkplek mee konden werken.

6. Proef op de som

6.1 Testscenario

Ervoor zorgen dat leveranciers tijdig de gewenste prototypes realiseren - en dat liefst in dezelfde periode om makkelijker te kunnen vergelijken - is nog maar een eerste stap. Minstens even belangrijk is het om een goed uitgewerkt testscenario te hebben, waarmee de prototypes op hun functionaliteit beoordeeld en vergeleken kunnen worden. Daarin kunnen onder meer de belangrijkste zoekcriteria worden getest met behulp van standaard zoekvragen. Of een Booleaanse combinatie goed wordt uitgevoerd en of bij zoeken in het titelveld inderdaad alleen in titels wordt gezocht, is betrekkelijk eenvoudig te controleren. Daarbij kan bovendien meestal worden volstaan met ja/nee-antwoorden. Beoordeling van de kwaliteit van de relevantievolgorde van de resultaten bij "best-match" zoekvragen ligt aanzienlijk lastiger. Bij dergelijke zoekvragen kan de testers worden gevraagd de relevantie in te schatten van de resultaatdocumenten op bepaalde rangnummers (bijvoorbeeld 1-20, 50, 100, 500) en tevens bij elk van die documenten aan te geven welke van de zoektermen hoe vaak in welk onderdeel (titel, samenvatting, volledige tekst) voorkomt. Zo kan getracht worden ook een meer objectieve indruk te krijgen van de factoren die de software bij het ranken van de zoekresultaten laat meespelen. Of een zo tijdrovende analyse altijd nodig is, hangt af van het belang dat gehecht wordt aan de kwaliteit van dit aspect van de zoekmachine.

Voor het uitvoeren van de test moeten voldoende zoekspecialisten uit de organisatie worden ingezet. Meestal zal het testen namelijk veel werk zijn; bovendien kan het op den duur wat saai worden dezelfde zoekvraag telkens in drie systemen te moeten uitvoeren en elke keer de resultaten daarvan te analyseren. In de praktijk blijkt dat gevaar van saaiheid wel mee te vallen; velen blijken het spannend te vinden om met dergelijke proefsystemen aan de gang te gaan. Bovendien krijgt men zo het gevoel aan de uiteindelijke keuze bij te dragen, waardoor ook al vroeg draagvlak voor acceptatie van het nieuwe systeem wordt gecreëerd.

Test van de prototypes bij de Universiteitsbibliotheek Utrecht

De test van de Utrechtse POC begon na uitgebreide voorbereiding door zowel de bibliotheek als de leveranciers. Hoewel twee maanden tevoren was afgesproken welke twee weken de testsystemen beschikbaar moesten zijn, was geen van de leveranciers op tijd klaar. Vooral het laden van 1 miljoen records viel hen tegen. Bij de bibliotheek maakte ondertussen een zoekspecialist testvragen op basis van de belangrijkste zoekcriteria uit het PVE. De testcoördinator probeerde de vragen uit op de oude zoekmachine die hiervoor met dezelfde 1 miljoen records was gevuld. Vragen die voor meerdere uitleg vatbaar bleken, konden nog worden aangepast.

De testcoördinator zorgde voor een vrij grote groep testers van uiteindelijk 15 mensen. Dat was in de eerste plaats om het werk te spreiden, maar bovendien kon zo al vroeg draagvlak voor het nieuwe systeem gecreëerd worden. Deze testers spelen namelijk ook een belangrijke rol om gebruik van het systeem door de klanten – studenten en medewerkers van de universiteit – te ondersteunen en te stimuleren.

>>

De testers werden in een bijeenkomst voorgelicht over het doel van de test, het soort testvragen en het antwoordformulier dat men moest gebruiken. Daarnaast werden de mensen aangemoedigd om na de test nog wat te 'spelen' met de zoekmachines, op te schrijven wat ze daarbij nog opviel, alsmede hun voorkeur uit te spreken. Van meet af aan was duidelijk dat ook de subjectieve beleving van de diverse systemen een rol zou spelen bij de uiteindelijke keuze.

De testers werd gevraagd binnen de testperiode van twee weken 24 uur te reserveren voor de test die ze op hun eigen werkplek zouden kunnen doen. In de praktijk bleken er aanzienlijke verschillen in werkelijk benodigde tijd per tester. Dat kon van de toebedeelde vragen afhangen, maar was ook individueel bepaald. Bij de evaluatie van de test bleek dat men in het vervolg liever in een apart leslokaal wilde testen, weg van de werkplek waar men dikwijls gestoord werd. Ook een direct aanspreekbare begeleider werd geprefereerd boven contact via email, zoals nu was geregeld.

Elke tester stelde zijn vragen aan alle drie zoekmachines; deze moesten dus in dezelfde periode voor de testers beschikbaar zijn. Om resultaten te kunnen vergelijken en er zeker van te zijn dat alle vragen minstens één keer echt goed gedaan zouden zijn, werd elke vraag door twee testers uitgevoerd. De testsystemen wilden namelijk nog wel eens haperen en niet iedereen was in de gelegenheid de vraag op een later moment nog eens over te doen. Bedoeling was ook dat testers met elkaar konden overleggen (iedereen had een overzicht hoe de vragen waren verdeeld), maar dat kwam niet van de grond.

Een grote steun bij het testen van relevantie bleek de (oude) Google toolbar in Firefox te zijn. Met behulp van de toolbar kon men zoektermen en woordstammen in zoekresultaten in afzonderlijke kleuren laten oplichten. Vooral het zoeken van woordstammen in abstracts had anders enorm veel tijd gekost. Dit was nodig om een oordeel over relevantie-ordening met kwantitatieve gegevens te kunnen correleren, maar zo werd ook gecontroleerd of de door het systeem afgeleide woordstammen niet teveel ongewenste woordvarianten opleverden.

Tijdens de testperiode werd met de testers gecommuniceerd via email en door het regelmatig rondsturen van het testdocument dat steeds werd aangevuld. Ook na de testperiode werden de testers op de hoogte gehouden van de voortgang en tijdens een evaluatiebijeenkomst werden de uitkomsten en resultaten van de test uitgebreid toegelicht terwijl men ondertussen van de borrel genoot.

Een kleine groep mensen bestaande uit de zoekspecialist, de testcoördinator en een ontwikkelaar hebben de resultaten van de test geanalyseerd. Dat kostte relatief veel tijd door de complexiteit van de relevantiebeoordelingen. Bovendien moesten vragen soms opnieuw gesteld worden, omdat een zoekmachine eerder gehaperd had of omdat een gegeven antwoord niet werd vertrouwd.

Omdat alle drie zoekmachines in principe voldeden aan de eisen uit het PvE, werd ook de subjectieve beleving van testers van belang voor de uiteindelijke keuze. Opvallend was dat de meesten een voorkeur hadden voor een zoekmachine met een relevantie-ordening die leek op die van de oorspronkelijk gebruikte en dikwijls bekritiseerde zoekmachine. De zoekmachine met een op andere principes gebaseerde relevantie-ordening werd door meerdere mensen expliciet afgewezen.

6.2 Technische criteria

Voor de technische eisen uit een PVE is een soortgelijke test middels een POC veel lastiger te realiseren. Niettemin dienen de ICT-medewerkers die verantwoordelijk worden voor verdere ontwikkeling en beheer van het systeem zich op zijn minst een goede indruk te kunnen vormen hoe "toegankelijk" de software voor hen is. Dit heeft betrekking op het installeren van de software, op het configureren daarvan voor de specifieke implementatie, op de daarvoor noodzakelijke aanpassingen van bijvoorbeeld de interfaces, op het koppelen met andere systemen, op het optimaliseren van het systeem voor performance, enzovoort. Dit kan onder meer door:

- bestuderen van technische documentatie van de leveranciers,
- gesprekken, uitleg en technische demonstraties door degenen die de POC's technisch realiseren,
- bezoek aan degenen die technisch beheer van de betreffende software uitvoeren bij andere gebruikers.

Op basis hiervan kunnen deze technici ook inschatten hoe makkelijk de systemen aan specifieke eigen wensen aangepast kunnen worden en hoe makkelijk noodzakelijke extra zaken ontwikkeld of 'bijgebouwd' kunnen worden. Tevens dient zo inzicht te worden gekregen in eventuele zwakheden van de daarvoor te gebruiken technieken en tussenlagen. De gesprekken met andere gebruikers kunnen bovendien een indruk geven van de kwaliteit van de ondersteuning die bij implementatie en latere gebruiksproblemen van de leverancier verwacht kan worden. Dit technische onderzoek kan er verder voor zorgen om bij voorbaat al draagvlak bij de technische afdeling te creëren.

7. Selectie

7.1 Uitwerking van de test

Wanneer een POC gerealiseerd is, vormen de resultaten van de daarin uitgevoerde functionele tests een belangrijk element voor de uiteindelijke keuze. In de eerste plaats wordt daarmee nog eens terdege gecontroleerd of de producten inderdaad voldoen aan (bijna) alle eisen uit het PVE, zoals was voorgespiegeld. Daarnaast kan bij bepaalde eisen ook worden geprobeerd de kwaliteit te bepalen, hoe goed daaraan wordt voldaan. Dat kan bijvoorbeeld het geval zijn bij de relevance ranking van de zoekresultaten.

Als geen volledig gecontroleerd corpus wordt gebruikt zoals bij TREC, de jaarlijkse competitie voor experimentele text retrieval systemen, blijkt de analyse van de resultaten van de daarvoor gestelde 'probabilistische' testvragen in de praktijk tamelijk complex. Daarbij speelt bijvoorbeeld de vraag mee of de relevantie moet worden beoordeeld op grond van vergelijking met een al bekende zoekmachine, op grond van de in 6.1 genoemde objectieve factoren of op de persoonlijke, per definitie meer subjectieve, maar wel meer realistische beleving van de testers. In het Utrechtse keuzeprocés heeft die laatste wijze van beoordeling uiteindelijk de doorslag gegeven, ook met het oog op het beoogde draagvlak. Eén zoekmachine werd door meerdere personen echt afgewezen, ook al zouden de zoekresultaten volgens een meer objectieve analyse gemiddeld niet significant slechter horen te zijn.

De resultaten uit de POC kunnen aanleiding zijn tot herziening van de oorspronkelijke beoordeling van bepaalde punten uit het PVE, zodat de score van de pakketten op het hele functionele deel van het PVE nog eens degelijker kan worden nagerekend. Ook kunnen nog eens specifiek scores worden bepaald voor de in de POC geteste aspecten – in principe immers de belangrijkste uit de functionele eisen.

7.2 Prijsopgave - Request for Quotation

Parallel aan de POC kan aan de betreffende leveranciers een zogenaamd "request for quotation" worden gevraagd. Dat is een veel gedetailleerder prijsopgave dan bij de RFI was gevraagd. Daarbij wordt dan preciezer rekening gehouden met de specifieke situatie bij de klant. Dat betreft dan onder meer:

- hoeveel (gelijktijdige) gebruikers er maximaal mogen of moeten zijn,
- op hoeveel servers of hoeveel processoren de software uiteindelijk gaat draaien,
- hoeveel afzonderlijke indexen of separate systemen ermee gebouwd (mogen) worden,
- hoeveel documenten er mee geïndexeerd moeten worden,
- welke extra modules eventueel nog benodigd zijn,
- welke op maat gemaakte aanvullingen vereist zijn,
- of het product inclusief de volledige implementatie geleverd moet worden,
- of er ook een onderhoudscontract bij hoort,
- of men daarbij ook automatisch recht heeft op updates,
- of ook training van medewerkers, support e.d. daarin betrokken zijn.

Ook hier is het zaak aandacht te besteden aan de vergelijkbaarheid van de uiteindelijk verkregen offertes. Soms moet dat in een iteratief proces gerealiseerd worden.

Dit is meestal ook het moment om al prijsonderhandelingen te beginnen. Bij dure producten blijkt er vaak nog wel wat van de prijs af te kunnen. Daar kunnen allerlei aanleidingen voor zijn. Van de prozaïsche overweging dat een account manager voor het afsluiten van het kwartaal nog een bepaalde omzet moet halen, tot de wens een bepaalde organisatie beslist als klant te willen hebben, als toonaangevend voorbeeld voor een hele categorie die tot dusverre in de klantenportefeuille ontbrak. Soms ook wordt geen werkelijke korting gegeven, maar wordt voor het oorspronkelijke bedrag een ruimere licentie geboden, worden extra modules gratis meegeleverd, of wordt gratis training aangeboden. Aanvankelijk duurder lijkende producten blijken zo toch weer concurrerend te kunnen worden met als veel goedkoper ingeschatte producten, al compliceert dit soms het streven de offertes goed vergelijkbaar te houden. Enige gezonde argwaan of de gegeven korting uiteindelijk niet toch weer op een andere manier wordt teruggepakt is overigens ook hier op zijn plaats.

7.3 Verdere afwegingen

Bij de uiteindelijke keuze zullen vrijwel altijd ook nog andere elementen een belangrijke rol spelen. Dat zijn

- De technische kwaliteiten van het systeem, zoals die tijdens de POC ook al zo goed mogelijk waren beoordeeld.
- De projectmatige realiseerbaarheid van een geheel aan de eigen eisen aangepast systeem. Daarbij onder meer de afweging hoeveel functionaliteit al standaard leverbaar is, hoeveel nog ontwikkeld moest worden, hoe makkelijk en hoe snel het ontwikkelen van extra functionaliteit en integratie met eigen modules naar verwachting gaat en welke support bij implementatie en verdere ontwikkeling te verwachten is.
- De kwaliteit en degelijkheid van de leverancier, zoals die in de betreffende eisen uit het PVE (zoals aangeduid in 3.7) voorkwamen.

De uiteindelijke afweging hoe belangrijk elk van de vijf hier genoemde factoren

- kwaliteit
- prijs
- techniek
- projectmatige realiseerbaarheid
- vertrouwen in de leverancier

bij de uiteindelijke beslissing moet meetellen, zal in elke situatie en bij elke organisatie verschillend liggen. Daarover vallen hier dus geen algemene adviezen te geven.

8. Ervaringen

Een zeer uitgebreide keuzeprocedure zoals hier beschreven, blijkt vaak een lange doorlooptijd te hebben. Dat kan in vrijwel alle fases spelen, zeker ook in die waar je van anderen afhankelijk bent.

Bij de Universiteitsbibliotheek Utrecht heeft het totale traject, van het bedenken van de procedure tot het tekenen van het contract, uiteindelijk vijftien maanden in beslag genomen. Het besluit bij welke drie leveranciers een POC zou worden aangevraagd lag ongeveer halverwege die periode. Dat het eerste deel van het traject al betrekkelijk veel tijd kostte, kwam deels doordat aanvankelijk een andere opzet was voorzien om van de longlist tot een shortlist te komen. Een aantal erkende zoekmachine-experts in binnen- en buitenland was gevraagd als een soort externe adviseurs een rijtje favorieten te noemen voor het type toepassing als de onze. Dit bleek in de praktijk echter niet van de grond te komen, zodat alsnog moest worden besloten een adviesbureau een volledig RFI te laten verzorgen. Ook het verzamelen en verwerken van die gegevens en het opstellen van het advies nam meer tijd dan verwacht. Vervolgens bleek ook de voorbereiding van de POC langer te duren en meer praktische problemen op te leveren dan - kennelijk ook door de betreffende leveranciers - was voorzien. Of twee maanden echt te kort was om de POC-systemen te realiseren, of dat men te laat was begonnen weten we natuurlijk niet. In elk geval leek men zich op het indexeren van een miljoen records te hebben verkeken, zodat alle drie testsystemen uiteindelijk pas een week later beschikbaar waren dan de met de testers afgesproken datum. Een dergelijke uitloop kan dus ook nog problemen opleveren met de afstemming en afspraken binnen de eigen organisatie.

Na de definitieve keuze, hoe zorgvuldig ook gedaan en hoe degelijk beargumenteerd, kan de implementatie van het systeem ook nog meer tijd kosten dan voorzien. Een van de redenen daarvoor in Utrecht was de grootte van de te doorzoeken collectie. Tussen een POC op 1 miljoen records en een werkelijk systeem met intussen meer dan 12 miljoen records bleek nog een wezenlijke technische grens te worden overschreden. Voor het specifieke soort veldgestructureerde gegevens waar het om ging, bleek het onderste uit de kan gehaald te moeten worden op het laagste niveau van hardware en operating system, om ook voor complexe zoekopdrachten nog snelle responsetijden te garanderen. De keuze van de hardware moest daar op het laatste moment nog aan worden aangepast. Omdat geen van de gesproken gebruikers in een exact vergelijkbare situatie had verkeerd, was deze problematiek niet naar voren gekomen uit de met hen gevoerde gesprekken.

Voor een eventuele volgende pakketselectie zien we, op grond van deze ervaringen, niettemin geen aanleiding het keuzetraject wezenlijk anders te laten lopen. Achteraf moesten we namelijk concluderen dat de in onze situatie ondervonden problemen ook niet makkelijk voorkomen hadden kunnen worden. Men moet zich dus realiseren dat zelfs een goed doorlopen keuzetraject met een vrij uitgebreide POC, nog geen garantie hoeft te zijn dat de uiteindelijke implementatie in een tamelijk complexe toepassing, geheel zonder problemen verloopt. Een dergelijk keuzetraject blijkt zich wel goed te lenen om vrij veel medewerkers als testers en beoordelaars een actieve rol te laten spelen, ten behoeve van het draagvlak voor een nieuw systeem.

Websites van genoemde organisaties:

[Delphi-group] <http://www.delphigroup.com/>

[Forrester] <http://www.forrester.com/>

[Gartner] <http://www.gartner.com/>

[Search Expertise Centrum] <http://www.searchexpertisecentrum.nl/>

[Searchtools] <http://www.searchtools.com/tools/tools.html>

[TREC] <http://trec.nist.gov/>

Literatuur:

[Quesenbery 2003] Quesenbery, W., Designing a search people can really use. *Intercom 2003 (12) 18-21*; <http://www.wqusability.com/articles/searh-intercom-200312.pdf>

[Resnick 2006] Resnick, M.L. en M.W. Vaughan, Best practices and future visions for search user interfaces. *Journal of the American Society for Information Science and Technology*, 57 (6) 781-787 (2006)

[Reynolds 2006]
<http://www.delphigroup.com/events/ii2006/presentations/reynolds.pdf>

[Rappoport 2005] Rappoport, A., Search user interface and user experience. *onderdeel van Searchtools website; geraadpleegd 27 september 2006*; <http://www.searchtools.com/info/user-interface.html>

[Sieverts & Teubner 2006] Sieverts, E.G. en M.A. Teubner, Van PVE via RFI langs POC en RFQ naar finish; een keuzetraject voor zoekmachinesoftware. *InformatieProfessional*, 10 (5) 28-33 (2006)