

Search & Retrieval; de Googlificatie van onze samenleving

Eric Sieverts

**Instituut voor Media en Informatie Management (HvA) &
Universiteitsbibliotheek Utrecht**

Er zijn al heel lang zoekmachines op internet. In 1994 was Lycos de eerste "echte" zoekmachine voor het web, waarmee toen amper 1,5 miljoen webpagina's doorzocht konden worden. In 1996 verscheen AltaVista als nieuwste, grootste en ook al snel als meest populaire. Aanvankelijk had ook AltaVista nog niet meer dan 30 miljoen doorzoekbare pagina's.



zoekmachines zijn er al lang

lycos
in 1994 de eerste "echte", met bijna 1,5 miljoen pagina's

altavista
in 1996 de nieuwste grootste, met ruim 30 miljoen pagina's

maar geen van alle heeft ons gedrag én zelfs onze taal zo beïnvloed als Google

Naast Lycos en AltaVista verschenen in die periode nog veel meer zoekmachines op het toneel, waarvan Excite, Infoseek, Northern-Light, Inktomi, Hotbot en AllTheWeb tot de bekendste behoorden. Toch heeft geen enkele zoekmachine uit die generatie zo'n impact gehad op ieders informatiegedrag en zelfs op ons taalgebruik, als - enkele jaren later - Google.

Wat is de oorzaak geweest van die veel grotere invloed van Google? Wat maakte Google zo anders dan al die eerdere zoekmachines? Daarbij hebben mijns inziens vooral de volgende punten meegespeeld:

- In de eerste plaats het kale interface, dat zoeken voor iedereen zo schijnbaar eenvoudig maakte. Je hoefde geen specialist meer te zijn om ermee te kunnen werken.
- Daarnaast ook zijn goede zoekresultaten. Google introduceerde een heel nieuwe manier van relevantieordening die in de praktijk veel beter bleek uit te pakken dan de methoden die tot dan toe door andere zoekmachines waren gebruikt. De eerst getoonde resultaten bleken inderdaad meestal de beste, waardoor gebruikers bleven terugkomen. Ook de grote dekking (intussen meer dan 20 miljard webpagina's en -documenten, bijna 1000x zo veel als AltaVista 10 jaar geleden) zorgde dat je vaker een goed resultaat kreeg. Juist voor simpele vragen van de grote massa bleek Google een prima hulpmiddel.
- Verder waren PR en timing een sterke kant van Google. Al lijken er nu wat barstjes te komen in het beeld van "aardig" bedrijf dat geen kwaad doet ("Don't do evil"), toch overheerst nog steeds het idee van "die leuke jongens van Google". Bovendien bleken ze de concurrentie meestal net een stapje voor te zijn bij de introductie van nieuwe snuffjes en nieuwe producten.

wat maakte Google anders ?

- zijn "kale" interface
- zijn goede zoekresultaten
(al geeft elke zoekmachine -bijna- exact wat je vraagt)
 - andere, betere relevantie-ordening
 - grote dekking (> 20 miljard) geeft vaker goed resultaat
 - goed voor simpele vragen van de grote massa
- zijn PR en zijn timing
 - company motto: "Don't do evil"
 - everyone loves Google (or don't we?)

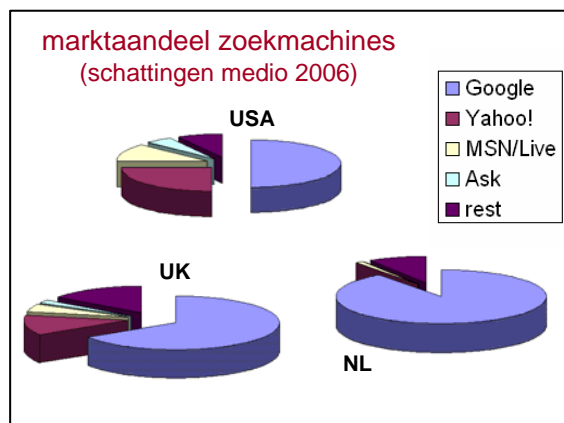
Eric Sieverts | e.g.sieverts@uu.nl | http://www.library.uu.nl/medew/eric | e.g.sieverts@hva.nl

het succes van Google

- zoeken is een "commodity" geworden
 - iedereen gebruikt thuis een zoekmachine
 - iedereen gebruikt op het werk een zoekmachine
 - iedereen gebruikt onderweg een zoekmachine?
- iedereen verwacht altijd overal te kunnen zoeken
 - "the ubiquitous search box"
- iedereen verwacht er altijd alles mee te vinden
 - "ambient findability"
- Google is daarbij de "maat der dingen" geworden
 - de usability benchmark?
 - de "Google experience"
- kortom: Google is synoniem met zoeken

Eric Sieverts | e.g.sieverts@uu.nl | http://www.library.uu.nl/medew/eric | e.g.sieverts@hva.nl

Het gevolg is dat zoeken in feite een "commodity" is geworden. Het is even gewoon als water en electra die "uit de muur komen". Iedereen is gewend geraakt overal een zoekmachine te (kunnen) gebruiken: thuis, op het werk en zelfs onderweg. Men spreekt wel van de "ubiquitous search box", het alom aanwezige zoekvenster, waarmee men verwacht altijd alles te kunnen vinden. Kijk maar eens naar twee populaire boeken: "Ambient findability" dat Peter Morville hierover schreef, en "The Search" van John Battelle, dat als ondertitel meekreeg "How Google and its rivals rewrote the rules of business and transformed our culture".







Daarbij is Google ook steeds meer de "maat der dingen" geworden, een soort "usability benchmark" voor ieder zoekstelsel. Als een zoekscherm er ingewikkelder uitziet dan Google, wordt het niet gebruikt. Google is letterlijk synoniem geworden met zoeken: iedereen Googlet zijn informatie bij elkaar. Opmerkelijk is overigens dat voor het zoeken op internet, het marktaandeel van Google in de VS nog lang niet zo hoog is als in

Nederland. In de VS heeft de concurrentie nog zo'n 40% marktaandeel, in Nederland niet eens meer 5%.

Search en retrieval in een professionele setting

Zoals in mijn inleiding beargumenteerd, heeft Google hooggespannen verwachtingen doen ontstaan ten aanzien van de "findability" van informatie. In die zin heeft Google ook voor Business Intelligence het zoeken op de kaart gezet. Dat betreft zowel het zoeken naar externe informatie, waarbij Google voor velen het eerst aangewezen hulpmiddel is, als het zoeken in de schat aan ongestructureerde informatie die binnen veel bedrijven aanwezig is. We moeten ons echter wel realiseren dat lang niet in elke situatie zo makkelijk aan die hooggespannen verwachtingen kan worden voldaan. Een intranet binnen een organisatie is nog heel iets anders dan het internet waarop Google naam heeft gemaakt. Methoden die op internet werken, zoals relevantie-ordening die vooral op linking-patronen gebaseerd is, hoeven helemaal niet zo goed te werken voor "enterprise search", het zoeken binnen de eigen informatiecollecties van organisaties. Bovendien zijn eisen die in een professionele omgeving worden gesteld aan relevantie en volledigheid van zoekresultaten, vaak heel anders - veel hoger - dan

in de consumentenomgeving thuis. Men zal zich bij een bedrijf veelal niet kunnen permitteren een belangrijk intern rapport te missen, terwijl het er op internet weinig toe doet dat ook nog ergens een niet gevonden 388^{ste} beschrijving van een bepaald hobby- onderwerp aanwezig is.

 <p style="text-align: center;">search & retrieval</p>  <p>door Google hooggespannen verwachtingen niet in elke situatie makkelijk daaraan te voldoen</p> <ul style="list-style-type: none">- een intranet is iets anders dan internet- wat op internet werkt (methode van relevance ranking) hoeft nog niet te werken op een intranet en voor "enterprise search"- eisen aan relevantie en volledigheid van zoekresultaat in werkomgeving anders dan in consumenten-omgeving <p><small>Eric Sieverts e.g.sieverts@uu.nl http://www.library.uu.nl/medew/eric e.g.sieverts@hva.nl</small></p>	 <p style="text-align: center;">"search" binnen organisaties</p>  <p>uit onderzoek van de Delphi Group (april 2006):</p> <ul style="list-style-type: none">- 34% van medewerkers zoekt > 6 uur/week- 42% van medewerkers besteedt > 40% van die tijd aan doorploegen van irrelevante informatie- 67% heeft geen zoekfunctie of heeft vaak vind-problemen; slechts 3% zegt: "intranet search is great"- 52% is ontevreden met "search experience" <p>daarom uitdaging voor elke organisaties te zorgen voor:</p> <ul style="list-style-type: none">- kwaliteit van zoekinterface en hele "user experience" <p><small>Eric Sieverts e.g.sieverts@uu.nl http://www.library.uu.nl/medew/eric e.g.sieverts@hva.nl</small></p>
--	--

Dat hier inderdaad een probleem ligt, blijkt uit onderzoek dat de Delphi-group regelmatig onder vooral Amerikaanse bedrijven houdt. Uit een onderzoek van vorig jaar bleek dat ongeveer 1/3 van de medewerkers van de onderzochte bedrijven meer dan 6 uur per week met zoeken bezig is. Maar ook dat 42% van die medewerkers meer dan 40% van die tijd kwijt raken aan het doorploegen van niet-relevante informatie. 67% heeft vaak "vind-problemen" of heeft helemaal geen zoekfunctie. 52% blijkt ronduit ontevreden over de "search experience" die wordt geboden. In bedrijfszoek-systemen is die Google-benchmark kennelijk nog niet vaak doorgedrongen. Voor veel organisaties vormt het verbeteren van de kwaliteit van de zoekfunctionaliteit en het zoekinterface van hun systemen dus een belangrijke uitdaging.

In dit kader is het goed eens te kijken naar zowel de sterke, als de – wel degelijk aanwezige – zwakke punten van Google. Dat kan ons wat leren voor de rol die zoekfunctionaliteit binnen bedrijfsomgevingen kan spelen, of dat nu bedoeld is voor het vinden van externe of van interne informatie.

Wat we kunnen leren van Google: 10 redenen waarom Google goed werkt

Ondanks de verschillen die er zijn tussen de eisen die gesteld worden aan zoeken binnen een bedrijfsomgeving en zoeken op internet, is het goed eens te kijken naar tien belangrijke punten waarop Google in elk geval goed scoort. Sommige daarvan zijn namelijk wel degelijk relevant bij het denken over een lokaal zoekstelsel dat ook voor BI nuttig zou kunnen zijn.

1. Google is (op internet) de grootste

Sinds ruim twee jaar geeft Google op zijn beginpagina zelf niet meer op hoeveel pagina's het doorzoekt en ook op sites met zoekinformatie, zoals Searchenginewatch, worden geen groottes van zoekmachines meer vermeld. Uit het vergelijken van zoekresultaten van verschillende zoekmachines blijkt echter wel dat Google vaker meer oplevert dan runner-up Yahoo!


Google is zelfs flink wat groter dan de zeker in Nederland wat minder bekende concurrentie, de zoekmachines Ask, LiveSearch, Exalead of Gigablast. Dat neemt overigens niet weg dat elk van die andere zoekmachines regelmatig informatie kan opleveren die niet in Google zit.

2. Google is zo eenvoudig in gebruik

Het “kale” zoekscherm van Google is de usability benchmark voor zoeksystemen geworden. Iedereen kan met Google uit de voeten en er iets mee vinden, zonder dat je eerst gebruikersinstructie nodig hebt. Systemen die ingewikkelder zijn (of alleen maar lijken) dan Google, dreigen niet meer gebruikt te worden. Maar wat gewone gebruikers vaak niet doorzien, is dat ook Google ingewikkelder kan zijn dan het lijkt! Als je de te gebruiken zoek-syntax maar kent, heb je allerlei geavanceerde zoekfunctionaliteit ter beschikking. Sommige informatie zit ook zo ingewikkeld in elkaar dat daarvoor ingewikkelder zoeksystemen nodig zijn. Feit blijft dat een eenvoudig ogend zoekscherm, dat automatisch zoekacties uitvoert op de wijze die het merendeel van de gebruiker - vaak onuitgesproken - bedoelt, eerder uitnodigt zo'n zoekstelsel te gebruiken.

3. Google heeft een uitstekend ranking mechanisme

Google's “pagerank” was het eerste mechanisme waarin iets als “democratisch bepaalde kwaliteit” van gevonden informatie meespeelde, een vroeg voorbeeld van gebruik van de “wisdom of the crowds”. Hoe meer links naar een site verwijzen hoe hoger die scoort. En hoe meer van die links afkomstig zijn van sites die zelf ook hoog scoren, hoe beter dat is. Dat houdt dus ook in dat het minder aan Google ligt, dan aan jouw pagina of website, als die niet altijd als eerste uit Google komen. Maar dat mechanisme werkt dus niet in een omgeving waar geen links zijn, zoals op een intranet of in een klassieker documentendatabase.

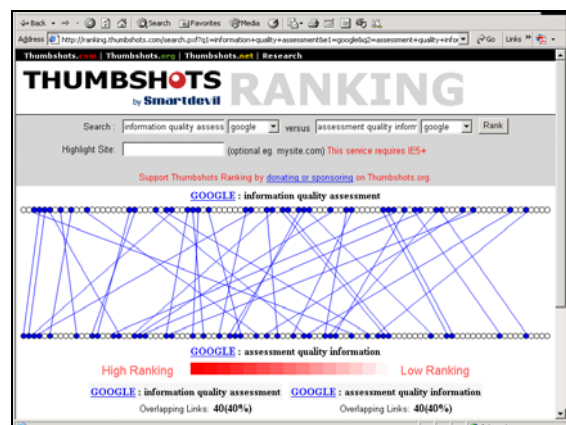


3: uitstekend ranking mechanisme



- “pagerank” was eerste mechanisme waarin “democratisch bepaalde kwaliteit” van gevonden informatie meespeelde
$$PR(x) = (1-d) + d \cdot \sum_y \{PR(y \rightarrow x) / C(y)\}$$
- dat *jouw* pagina niet altijd als eerste uit Google komt, ligt minder aan Google dan aan jouw pagina
- pagerank vooral bepalend voor volgorde bij “one-word queries”
maar: werkt niet als er geen links zijn (*intranetten*)!
- bij “intelligentere” queries ranking vooral op basis van andere parameters (*waar zoekwoorden voorkomen, woordnabijheid, woordvolgorde - phrase*)

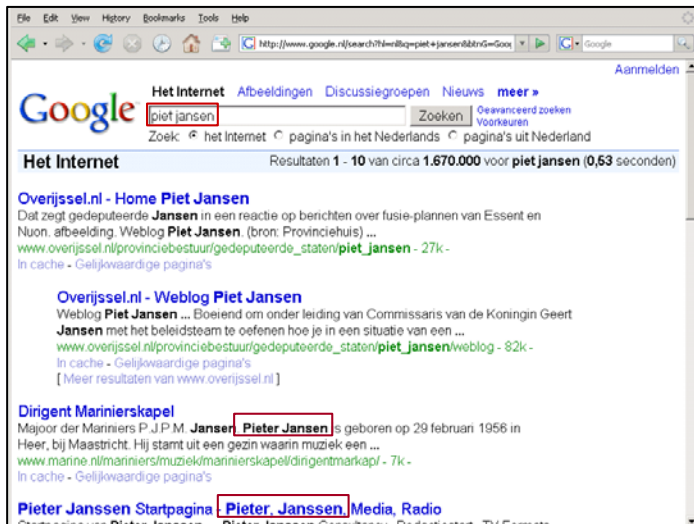
Eric Sieverts | e.g.sieverts@uu.nl | http://www.library.uu.nl/medew/eric | e.g.sieverts@hva.nl



Overigens is die pagerank vooral bepalend voor de presentatievolgorde bij eenvoudige “one-word queries”. Bij “intelligentere” zoekvragen wordt de ranking vooral bepaald op basis van andere parameters, zoals de plaats waar de zoekwoorden in een gevonden document voorkomen, hoe ze ten opzichte van elkaar in de tekst voorkomen (hun volgorde, hun onderlinge afstand, of ze een vast zinnetje zijn) enz. Dat zijn natuurlijk wel factoren die ook in andere omgevingen dan het internet gebruikt kunnen worden (en ook in de meeste andere moderne zoeksoftware worden toegepast).

4. Google biedt (soms) slimme automatische vraagverbetering

Of je een Engels zoekwoord nu in enkelvoud of in meervoud intikt, Google zoekt automatisch op zowel enkel- als meervoud en ook nog op enkele standaard uitgangen (althans op de Engelstalige Google-site). Google doet dat (op de Nederlandstalige site) ook met sommige Nederlandse woorden, maar ook weer niet heel consistent. Dat draagt er in elk geval toe bij dat zoekers een vollediger resultaat krijgen, ook als ze zelf niet bedenken dat de begrippen waarop ze zoeken in de te vinden teksten zowel in enkelvoud als in meervoud of als andere woordvormen kunnen voorkomen.

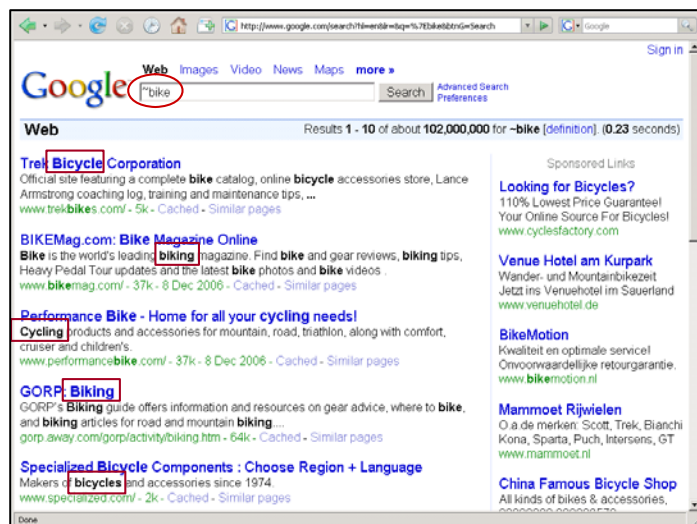


Google doet dat ook slim voor bepaalde vaste afkortingen, zodat “jfk” ook “john f kennedy” oplevert en “wwii” ook “world war II”). Op de Nederlandstalige site doet Google dat ook voor bepaalde Nederlandse namen (“Piet” vindt ook “Pieter” en “Jansen” ook “Janssen”). Alleen is niet erg duidelijk wat wanneer wel of niet gebeurt.

5. Google kan synoniemen aan een zoekvraag toevoegen

Door voor een (Engelse) zoekterm een ~ te zetten, zoekt Google ook op (Engelse) synoniemen van dat woord. Ook dit is erop gericht dat mensen niet meer zelf hoeven te bedenken dat het begrip waarnaar ze zoeken, ook wel eens met andere woorden omschreven kan zijn.

In de praktijk schiet Google hierbij zijn doel echter wat voorbij, doordat het begrip “synoniem” vaak erg ruim wordt opgevat. Daardoor levert het vaak meer rommel op dan dat het zoekresultaat erdoor verbetert. Of je bij het zoeken op “database” wilt dat ook wordt gezocht naar “access” (niet alleen de naam van een software-product, maar ook het zeer algemene Engelse woord “toegang”), is immers sterk de vraag.



Bovendien is dit een tamelijk onbekend stukje zoeksyntax dat de gebruiker heel bewust moet toepassen; hij/zij moet zich er dus al van bewust zijn dat synoniemen een zoekprobleem kunnen vormen.

6. Google biedt allerlei extra online tools

Via het gewone zoekvenster kun je Google ook allerlei andere dingen laten doen dan zoeken; vooral veel rekenwerk. Zo is Google:

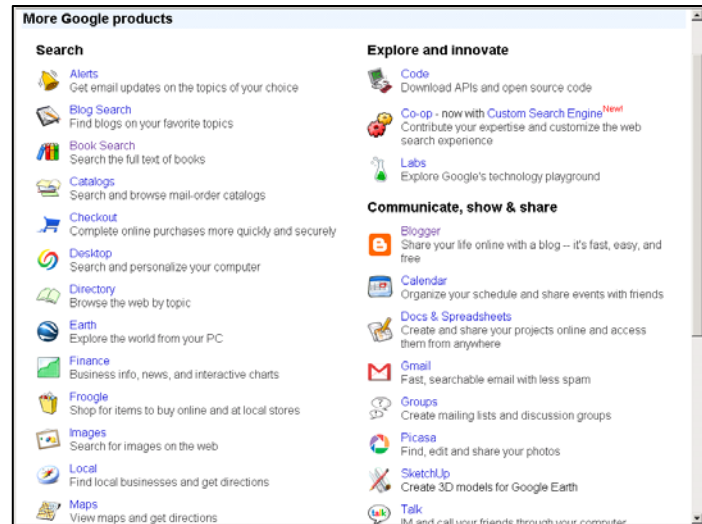
- rekenmachine (3*7/5)
 - omrekenaar (87F in C)
 - valutahulp (27USD in EUR)
 - definitiezoeker (define:relevance)
 - adreszoeker (650 Madison Avenue, Albany, NY)
 - telefoonboek (john smith, schenectady, ny)
- enz.,

waarbij tussen haakjes voorbeelden van wat je kunt intikken. Die adressen en telefoonnummers beperken zich overigens (nog) wel tot de VS. Via het interface dat je toch al voortdurend gebruikt, ook allerlei andere taken kunnen uitvoeren, is natuurlijk zeer aantrekkelijk.

7. Google biedt veel offline tools & online diensten

Google biedt allerlei gespecialiseerde diensten. Soms zijn dit apart op je PC te installeren programma's, soms zijn het webdiensten, soms zijn het combinaties daarvan en soms zijn het programmatjes die in je web-browser draaien. Een paar voorbeelden:

- Google toolbar die in combinatie met je web-browser allerlei handigheidjes biedt bij internet-gebruik;
 - Google desktop: zoekprogramma dat de hele inhoud van je eigen harde schijf full-text doorzoekbaar kan maken;
 - Picasa: voor het beheren en uitwisselen van foto's; het biedt een combinatie van lokaal beheren en online beschikbaar stellen;
 - Google mail: mailprogramma via het web;
 - RSS reader: om online via het web RSS-feeds te beheren en te lezen;
 - Google Docs & Spreadsheets: tekstverwerker en spreadsheet die in je browser draaien;
 - Blogger: host om weblogs te publiceren;
 - Googlepages: host om je eigen website te bouwen;
 - Google-coop: custom search engine op eigen selectie van te doorzoeken websites;
 - Google suggest: komt dynamisch met woordsuggesties terwijl je zoekwoorden aan het intikken bent;
- enz.



8. Google biedt additionele zoekmachines voor andere media

Behalve de zoekmachine voor gewone webpagina's biedt Google onder meer:

- image search (afbeeldingen)
 - newsgroup search (berichten uit nieuwsgroepen e.d.)
 - video search
 - blog search (op weblogs verschenen posts)
 - news search (berichten uit nieuwsbronnen, apart voor 10 talen)
 - book search (full-text gescande boeken)
 - google scholar (wetenschappelijke artikelen)
 - google maps / google earth
 - shopping search
 - finance search
 - desktop search
 - program code search
- enz.

Overigens heeft Google sinds kort een aantal van deze zoekdiensten in zijn gewone zoekmachine geïntegreerd, onder de naam “universal search”. Of het altijd nuttig is om ook nog plaatjes, video’s en nieuwsberichten gepresenteerd te krijgen, als je alleen naar een bepaalde webpagina op zoek was, kun je natuurlijk betwijfelen. In de praktijk valt het gelukkig erg mee hoe (weinig) vaak dat gebeurt. Anderzijds is het automatisch samenvoegen van resultaten uit verschillende bronnen ook een duidelijke trend in de professionele zoekwereld van de betaalde informatiebronnen, daar onder de benaming “content-integratie”.

9. Google biedt automatische attenderingsdienst

Bij Google kun je zoekvragen opslaan, om regelmatig per mail geattendeerd te worden, wanneer nieuwe informatie is verschenen die aan die zoekvragen voldoet.

9: persoonlijke attenderingsdienst

- regelmatige attendering via mail, op basis van eigen zoekvraag, uit:
 - web
 - nieuws
 - nieuwsgroepen
 - blogs

Eric Sieverts | e.g.sieverts@uu.nl | http://www.library.uu.nl/medewit/eric | e.g.sieverts@hva.nl

Daarbij kan het gaan om de aanwezigheid van jouw combinatie(s) van zoekwoorden

- in web-pagina’s,
- in het nieuws op Google-News,
- in berichten op nieuwsgroepen,
- in posts op weblogs.

Een bijzonder handig hulpmiddel, al blijkt het zeker niet altijd 100% foutloos te werken.

10. Google bewaart desgewenst je zoekgeschiedenis

Als je een persoonlijk account hebt aangemaakt, kan je zoekgeschiedenis door Google worden bewaard. Dat betreft dan niet alleen je zoekvragen, maar ook informatie over welke resultaten je metterdaad hebt aangeklikt.

Zo kun je terugvinden

- hoe je eerder hebt gezocht,
- wanneer je dat gedaan had,
- welke resultaten je eerder had bekeken (als je die wel hebt aangeklikt, maar hebt vergeten te bookmarken).

Hoewel Google in een privacy-statement meldt dat deze soms potentieel wat gevoelige informatie niet aan derden beschikbaar wordt gesteld, worden de laatste tijd steeds meer zorgen geuit of hier toch geen misbruik van gemaakt zal worden.

Wat we kunnen leren van Google: 10 redenen waarom Google (soms) geen goed idee is.

1. Met Google vind je niet wat je zoekt

In 40 miljard items op internet vind je (met Google) altijd wel iets dat een antwoord op je vraag is, maar vaak niet precies dat ene document waarnaar je op zoek was en waarvan je wist dat het bestond. In een lokaal systeem met 20.000 items is dat vaak nog veel sterker. Dat kan komen doordat

- het gezochte document toch net niet dat ene woord of die combinatie van woorden bleek te bevatten, waarop jij zocht,
- het begrip waarnaar je zocht in dat document alleen als werkwoordsvorm bleek voor te komen, terwijl jij op een zelfstandig naamwoord zocht,
- het woord waarop je zocht toch anders gespeld bleek te zijn,
-

Zeker voor een lokaal systeem zal dus een slimmere zoekmachine nodig zijn.

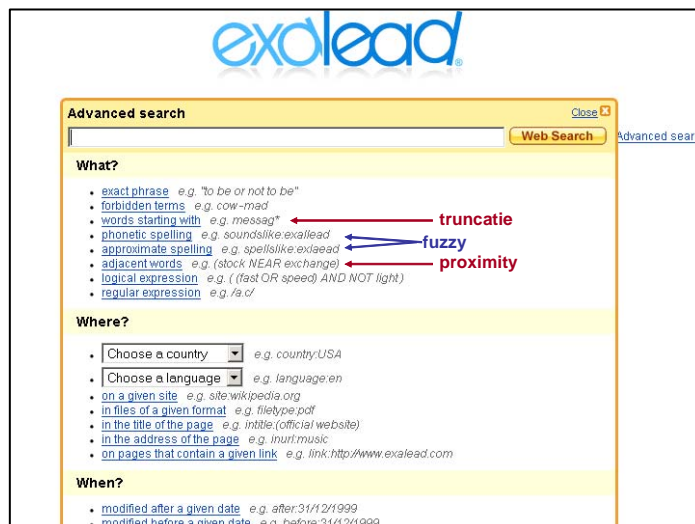
2. Met Google vind je ook wat je niet zoekt

Als elk woord uit de volledige tekst van documenten zoekbaar is, vind je vaak ook documenten waarin je zoekterm eigenlijk helemaal niet van belang is. Dit ondanks de methoden van relevantie-ordening die worden toegepast.

Als je Google-technologie ook toepast op je lokale materiaal, bijvoorbeeld met Google Desktop, maar je je spullen niet goed op orde hebt, dan vind je voortdurend ook allerlei ongewenste oude versies van documenten. Als je helemaal ongestructureerd zoekt, zul je, wanneer je bijvoorbeeld naar “meneer” Bakker zoekt, bovendien ook documenten over brood vinden. Dat betekent dat voor een lokaal systeem een slimmere en meer gestructureerde zoekmachine nodig is.

3. In Google ontbreekt allerlei zoekfunctionaliteit

Niet alleen in het kader van de eenvoud van het zoekinterface, maar ook om, ondanks de enorme hoeveelheid content en de gigantische aantallen gelijktijdige gebruikers, toch snelle responsetijden te kunnen garanderen, wordt allerlei wat ingewikkelder zoekfunctionaliteit niet aangeboden. Zo is er bijvoorbeeld geen mogelijkheid om zoektermen te trunceren (gebruik van wildcards) en kun je ook niet zoeken op combinaties van woorden die in de te vinden teksten beslist in elkaars nabijheid moeten voorkomen. Daar kan natuurlijk tegenin gebracht worden dat Google (in het



Engels) toch al automatisch op enkel- en meervoud zoekt, zodat truncatie nauwelijks meer nodig lijkt. En woordnabijheid telt bij de relevantie-ordening toch ook al flink mee. Bovendien bieden de meeste andere webzoekmachines deze functionaliteiten ook niet, met als enige uitzondering Exalead. Voor een lokaal systeem is het echter zeker nuttig om software te kiezen die dit soort dingen wel kan.

4. Google biedt geen goede tijdinperking

In de eerste plaats biedt Google een heel beperkte keuze van tijdperiodes waarop je een zoekresultaat kun tinperken; alleen:

- laatste 3 maanden
- laatste 6 maanden
- laatste 12 maanden.

Bovendien blijkt dat in de praktijk vaak niet goed te werken. Datzelfde geldt voor de ongedocumenteerde "daterange" optie die gebruik maakt van een juliaans datum format. Een mogelijkheid om heel gedetailleerd op datum te zoeken, ook op intervallen met per dag instelbare begin- en einddatum, en waarbij dat ook nog redelijk betrouwbaar gebeurt, is wel aanwezig bij AllTheWeb, AltaVista en Exalead. Voor het zoeken in eigen informatie is dit ook een volstrekt essentiële functionaliteit.

5. Google zoekt niet op metadata

Veel webpagina's bevatten in hun broncode zogenaamde metadata; bijvoorbeeld:

```
<meta name="keyword" content="business_intelligence, corporate_search, google">
```

Google doorzoekt deze metadata niet, omdat men vindt dat bouwers van websites te vaak misbruik maken van deze terfwoorden, met het doel "oneigenlijk" gevonden te worden. Toch is het toevoegen van metadata een essentiële methode om pagina's met weinig "eigen" tekst toch nog vindbaar te maken op daarvoor relevante woorden.


Andere zoekmachines doen dat daarom wel, zoals Yahoo, AllTheWeb, AltaVista, Ask, Hotbot en Gigablast, al indexeren die meestal niet meer dan alleen de eerste 16 of 24 keywords die zijn toegekend.

Zeker in interne systemen wil je ook op metadata kunnen zoeken. Bijvoorbeeld om bepaalde soorten documenten gestandaardiseerd terug te kunnen vinden, of om belangrijke documenten op bepaalde zoekwoorden voorrang te geven in de relevantie- ordening.

6. Google biedt geen reproduceerbare resultaten

Voor dezelfde zoekvraag blijkt Google vaak telkens verschillende aantallen resultaten te geven, ook als twee mensen tegelijkertijd dezelfde zoekactie doen. Wanneer je de resultaten bekijkt van een zoekactie die een afzienbaar aantal resultaten opleverde, zie je pas bij het bekijken van de hele lijst hoeveel echt is gevonden. En dat is meestal minder dan aanvankelijk gemeld.

Ook is meestal niet duidelijk wanneer op woordvarianten wordt gezocht. De resultaten van Booleaanse combinaties kloppen (daardoor?) vaak ook niet. Bovendien blijken er vaak verschillen tussen de resultaten uit de Nederlandse en de Amerikaanse versie van Google. Daarbij moet overigens gezegd worden dat andere webzoekmachines, zoals bijvoorbeeld Ask ook wel eens rare effecten opleveren.



6: geen betrouwbare aantallen



Google

recept doornhaai

Hot Internet

Resultaten 1 - 10 van circa 123 voor **recept doornhaai** (0,04 seconden)

Lekkere bereidingen en **recepten** met **hondshaai** (Vis)

1000 **recepten** op basis van vlees, vis, geogelte of wild en talijke ... In de handel zal men de **doornhaai** en de hondshaai meestal in moten aantreffen. ...

Unlabeled Document

Doornhaai staat ook wel bekend als zaagpaling. Dit is echter niet correct. ... Hoofdgerecht voor 4 personen. 600 gram **doornhaai**, 30 gram ui en in ringen, ...

Big Bang Revisited

recept doornhaai, 30, 17, 6, 13, 13, 9, 9, 6: "hogeschool van amsterdam", 32000, 20100, 4470, 27400, 20100, 5748, 5666, 11701 ...

voorbeeld:

- recept AND doornhaai 123 (in resultaat ook "recepten")
- recepten AND doornhaai 195 (in resultaat geen "recept")
- (recept OR recepten) AND doornhaai 123

Eric Sieverts | e.g.sieverts@uu.nl | http://www.library.uu.nl/medewit/eric | e.g.sieverts@hva.nl

7. Google levert veel minder backlinks

Een zoekoptie die veel webzoekmachines bieden, is dat je kunt zoeken naar alle webpagina's die een link bevatten naar een pagina waarvan je het URL opgeeft. Bij dat "link zoeken" vindt Google altijd vele malen minder webpagina's dan Yahoo, AllTheWeb, AltaVista of Exalead. Maar Google beweert wel dat het de minst belangrijke zijn (die met laagste pagerank) die ontbreken. Voor wie een volledig overzicht wil hebben is dit echter onaanvaardbaar.

8. Google biedt geen hulp bij het verfijnen van zoekvragen

Veel gebruikers zoeken – vanuit het oogpunt van zoekspecialisten – tamelijk slecht, waardoor ze vaak veel te veel vinden. Het zal dus vaak nodig zijn om een zoekvraag verder te verfijnen. Google biedt daar verder geen hulp bij.


Sommige andere zoekmachines doen dat wel, meestal door middel van technieken die gebaseerd zijn op statistische analyse van de woorden die in het aanvankelijke zoekresultaat voorkomen. Zo bieden onder meer Ask en Quintura (waar Yahoo achter zit) rijtjes begrippen die vermoedelijk verschillende contexten van het onderwerp representeren. Door op zo'n begrip te klikken, wordt de zoekactie met die term ingeperkt. Een metazoekmachine als Clusty deelt het zoekresultaat op dezelfde manier in een beperkt aantal clusters op.

 **8: geen hulp bij verfijnen zoekvraag** 

- omdat men vaak slecht zoekt en veel te veel vindt, is het vaak nodig zoekvraag te verfijnen
- sommige zoekmachines bieden daarbij hulp door statistische analyse van woorden uit zoekresultaat (Ask, Quintura/Yahoo, Clusty, ...)
- sommige zoekmachines delen resultaat (ook) op naar meer formele kenmerken - "parametric search" (Exalead)

Eric Sleverts | e.g.sleverts@uu.nl | http://www.library.uu.nl/medow/eric | e.g.sleverts@hva.nl

 **Web Search** Showing results 1-10 of 3,103,000

Search Tools: Look up 'maxima' in Encyclopedia | Dictionary | Thesaurus

Maxima
Homepage of the Maxima Project. Maxima is a system for the manipulation of symbolic and numerical expressions, including ...

Maxima - Download
Downloads for Maxima. For an installation from source code, you want maxima.x.y.z.tar.gz or maxima.x.y.z.src.rpm. But you know that ...

MaximMag
Online version of a men's magazine. Girls, entertainment, sports and stupid jokes.


Maxima.org :: News
The Home for Maxima Enthusiasts. Maxima of the Month: February 2006 Maxima of the Month: Car 1995 Arctic Pearl White (GHO) Maxima ...

Nissan Motor Co., Ltd.
Official site for Nissan cars, trucks and SUV's. Search, compare, view photos, locate a dealer, and request car quotes for your favorite 2006 ...


Narrow Your Search

- Nissan Maxima
- Nissan Maxima 2004
- 2002 Nissan Maxima
- Performance Nissan Maxima
- 1995 Nissan Maxima
- Princess Maxima
- Nissan Maxima Parts
- Maxima Club
- Maxima Fishing Line
- Van Maxima
- Zorregueta, Maxima
- 1997 Maxima
- Maxima FM
- Maxima Definition
- Turbo for Maxima
- Nissan Maxima

 **New search**



See results

 **Stones** Top 189 results of at least 57,982,800 retrieved for the query stones (definition) (details)

clusters sources sites

- All Results (110)
- Rolling Stones (7)
- Jewelry (11)
- Stepping Stones (8)
- Stones World (1)
- Sculpture (1)
- Cut (1)
- Kidney Stones (1)
- Granite (1)
- Standing Stones (1)
- Natural Stones (1)
- more | all clusters

Top News Find more news stories

Eye-Popping Gemstones
Loose Or In Custom Jewelry At Wholesale Prices Every Day - www.jewelz.com

Beads Wholesale Supply
Discounted Faceted Stone Beads Wholesale prices, retail webstore - www.stoneagehardware.com

1. Stone is a rock. See also stone skipping, curling. As a verb, to stone, a method of execution using rocks, stoning. A gemstone, as used in jewelry. A stone is a unit of weight equal to fourteen pounds. The hard covering enclosing the seed of a drupe such as a peach. As a verb, the intoxicating effects of cannabis, as in to stone. Stone, a Fresh heavy metal band. A calculus. "The Stones" is an early album by Psychedelic Percussion featuring the Moog synthesizer. en.wikipedia.org/wiki/Stone - [cache] - Wikipedia

2. Stoning is a form of capital punishment in which the convicted criminal is put to death by having stones thrown at them, generally by a crowd. In some cultures, this was seen as

 **Web Search** Results 1-10 of about 10 vendor hyperfine silicon

Did you mean: vendor hyperfine silicon

Related terms

- Double resonance exclude
- Electron spin exclude
- Quantum wells exclude
- Silicon carbide exclude

Languages

- English (67) exclude
- French (3) exclude
- German (1) exclude

File types

- Acrobat (pdf) exclude (2)
- Word (doc) (9) exclude
- Text (txt) (9) exclude

Geographic location

- North America exclude (6)
- United States exclude
- Canada exclude
- Europe (9) exclude
- Germany exclude
- Great Britain exclude
- Italy exclude
- Asia (9) exclude
- Russian Fed exclude

Search within results

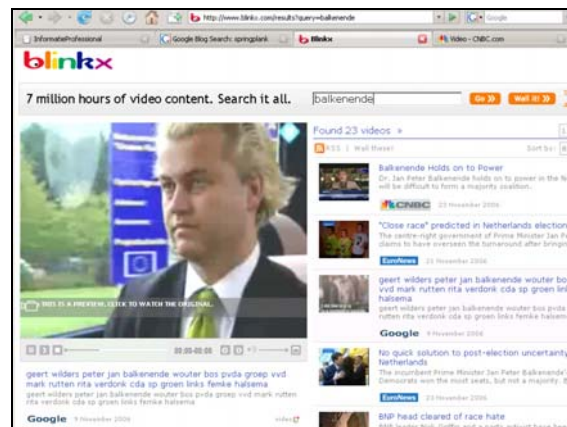
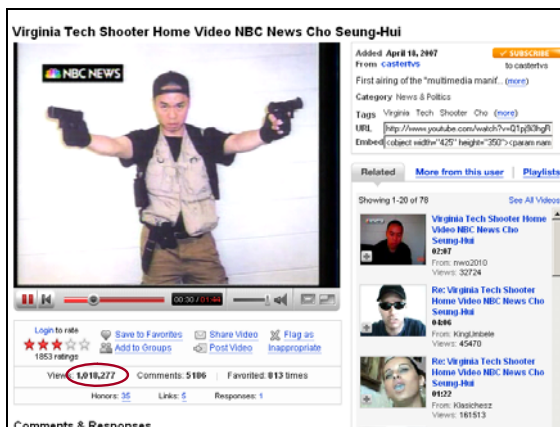
Een enkele zoekmachine (Exalead) deelt het resultaat ook op naar meer formele kenmerken, zoals taal, internet-domein of documentsoort, waarbij je vooraf al te zien krijgt hoeveel er bij die inperkingen overblijft. Zeker dit soort opdelingen is voor lokale zoeksystemen van groot belang. Daar wordt die zoekmethode veelal “parametric” of “faceted” search genoemd.

9. Voor andere media zijn er betere alternatieven

Google is destijds als één van de eersten met een zoekstelsel voor video begonnen. Daarbij richtte men zich aanvankelijk op publieke TV-programma’s (in Californië). De tekst van de ondertiteling voor slecht-horenden werd dan gebruikt om op te zoeken. Na enige tijd volgde een halfslachtige switch naar het uploaden van persoonlijke video’s door gebruikers. Intussen was YouTube daarvoor echter al veel populairder, waarbij vooral het uitwisselen en het via tags karakteriseren van het videomateriaal een grote rol speelde.

Net als bij Google is dit materiaal alleen doorzoekbaar op die door de makers meegegeven tags (trefwoorden) of korte beschrijvingen. Door YouTube op te kopen heeft Google daar vervolgens wel een zakelijke en inhoudelijke vinger in de pap gekregen.

Anderzijds is Blinkx een video-zoekmachine die ook professioneel materiaal bevat, van o.a. nieuwsdiensten. Door gebruik van (Engelstalige) spraakherkenning zijn veel video’s daar wel “full-text” doorzoekbaar op de gesproken tekst.



Enkele verschillen tussen beide systemen zijn hieronder weergegeven:

<u>YouTube</u>		<u>Blinkx</u>
uploaden	↔	webcrawler (robot)
kwaliteit vaak slecht (van mobieltjes)	↔	veel professioneel materiaal
zoekt op “metadata” (tagging)	↔	zoekt full-text (spraakherkenning)
meer gericht op browsen (“most popular”)	↔	meer gericht op zoeken

Aanvankelijk opgezet als zoekmachines voor podcasts (audio & video van omroepen, amateurs en ook bedrijven), doorzoekt EveryZING nu ook een deel van het videomateriaal van YouTube. Daarbij wordt dankzij spraakherkenning ook “full-text” gezocht. In de op het scherm getoonde tekst kun je dan zien waar jouw zoekwoorden voorkwamen.

10. Google is niet erg "web-2.0 - aware"

Veel van Google's slimme zoekoplossingen berusten op software en op geautomatiseerde analyse van grote hoeveelheden digitaal materiaal. Een kenmerk van veel aspecten van Web 2.0 toepassingen is juist dat het vooral "peopleware" is, dat de input (en inspanning) van mensen wordt gebruikt. Allerlei nieuwe diensten en startups leggen daarmee wel meer "de macht bij de gebruiker". En ook Yahoo! is veel sterker dan Google op dit soort oplossingen gericht.

In een paar steekwoorden wat karakteristieken van Web 2.0 met daarachter voorbeelden waar dat wordt toegepast:

- zelf publiceren (blogs, wiki's, foto's, video's, podcasts, ...),
- zelf het nieuws bepalen (newsvine, digg, postgenomic, ...),
- zelf bookmarken (del.icio.us, myweb, furl, connotea, ...),
- zelf taggen (overal: flickr, del.icio.us, digg, technorati, ...),
- zelf netwerken (hyves, myspace, orkut, facebook, linked-in...),
- zelf zoekmachines maken (rollyo, wink, yoono, google-coop),
- samenwerken en delen (overal).

Bij de diensten die Google zelf aanbiedt vinden we daar betrekkelijk weinig van terug. Toch speelt ook Google daar wel een rol, door onder meer:

- overnames van bedrijven die web 2.0 successen boeken,
- ontwikkelen van toepassingen die zijn gebaseerd op ajax-technologie in de browser (de eerder genoemde Google Docs & Spreadsheet)
- de open API voor mash-ups met Google-Earth
- ondersteuning van rss en blogging

In dit kader is het ook interessant eens te kijken in hoeverre die "social software" en dat "taggen" bij Web 2.0 toepassingen, concurrentie vormt voor het "echte" zoeken. Hieronder enigszins schematisch de verschillen tussen de Web 2.0 benadering en het "professionele" zoeken:

<u>Web 2.0</u>		<u>professionele zoeksystemen</u>
"most popular"	↔	zelf omschreven behoefte
tags	↔	metadata
tag cloud	↔	zoekvenster
andermans advies	↔	zelf zoeken
notification (rss)	↔	zelf zoeken
amusement	↔	werk en studie
Flickr.com	↔	klassieke image-search
YouTube	↔	Blinkx

Enkele conclusies en trends

Google is het bedrijf dat "search" voor iedereen en voor elk soort gebruik op de kaart heeft gezet. Met de daarmee gevestigde naam heeft het ook andere gerelateerde diensten zichtbaar gemaakt. Veel nieuwe technologieën sluiten aan op dat "search" paradigma. Anderzijds leggen veel recente Web 2.0 diensten meer nadruk op bladeren en het volgen van adviezen uit de eigen community, dan op gerichte eigen zoekacties. In beperkte mate geldt dat zelfs voor bepaalde professionele toepassingen. Ondanks de rol die Google heeft gespeeld voor de populariteit van "zoeken", betekent dat niet dat Google automatisch de beste is voor elke toepassing. Toch heeft de

concurrentie het op de consumenten-markt moeilijk, omdat Google juist daar de grote naam is. Voor enterprise search en Business Intelligence is Google zeker niet automatisch de beste, al moet gezegd worden dat lang niet elke manager zich realiseert dat "enterprise search" nog heel iets anders is dan "internet search".

Voorts dient men zich te realiseren dat "search" ook pas het eerste puzzle-stukje is, om de bij organisaties aanwezige ongestructureerde informatie ook voor Business Intelligence uit te baten. "Zoeken" is bijvoorbeeld nog lang geen text-mining: je vindt op zijn best waar je toch al bewust naar zoekt; nog onbekende trends en samenhangen komen er niet "vanzelf" uit.

	conclusies & trends	
<ul style="list-style-type: none">• Google zette "search" op de kaart en maakte andere gerelateerde diensten zichtbaar• nieuwe technologieën sluiten aan op het "search" paradigma• Google niet automatisch de beste voor elke toepassing (<i>toch heeft concurrentie het op consumenten-markt moeilijk</i>)• Google zeker niet de beste voor enterprise search en BI• enterprise search is heel wat anders dan internet search• search is pas eerste puzzle-stukje om ongestructureerde informatie in BI uit te baten (<i>het is nog geen text-mining</i>)• trend naar integratie van diensten en technieken• trend naar verdergaande personalisatie (<i>ook voor ranking</i>) <p><i>disclaimer: "mijn woorden geven geen garantie voor de toekomst"</i></p>		