



# masterclass “resources” op internet

Eric Sieverts

GO, 21 oktober 2008, Den Haag

# doelstelling / leerdoel

---

leerdoel van deze cursus:

- U bent in staat de voor uw onderzoek meest geschikte *bronnen* te selecteren, te karakteriseren en te gebruiken.
- U kunt anderen adviseren bij de keuze van informatie *bronnen*

# gedachtenbepaling vooraf

---



- wat is een **bron**?
- is **Google** een bron?
- is het web een bron?
- **NEE**  
*(althans voor deze cursus)*
- **Google** is een zoekingang op heel veel bronnen
- het web is een medium waarop heel veel soorten bronnen beschikbaar zijn

# gedachtenbepaling vooraf

---



uitgangspunt voor de cursus:

- een bron is een collectie van een bepaald soort informatie die online beschikbaar is,
- waarbij de soort bepaald kan zijn door de vorm (krantenartikelen, blogs, video's, plaatjes, ...),  
en/of door de inhoud (medische informatie, nieuws, wetenschap, ...)

# gedachtenbepaling vooraf



om "bronnen" te kunnen raadplegen maken we gebruik van *zoeksystemen* of *zoekingen*.

- in die zin is  een zoekingang op een veelheid aan bronnen.
- in die zin is  een zoekingang op weblogs.
- in die zin is  een zoekingang op wetenschappelijke bronnen.

# programma

---

- 10:00 / welkom, introductie
  - (0) inleiding, bronnen, zoekingen
- ca. 11:00 / thee- en koffiepauze
  - (1) boeken, (wetenschappelijke) artikelen
  - (2) feiten, naslag
  - (3) nieuws, kranten, tijdschriften
- ca. 12:30 / lunchpauze
  - (4) weblogs, feeds
  - (5) multimedia
  - (6) “oud” nieuws
- ca. 15:00 / thee- en koffiepauze
  - (7) algemene versus gespecialiseerde zoekingen
  - (8) beoordelen van bronnen
  - (9) bijblijven

# bekend verondersteld

---

- verschil tussen soorten diensten op internet
- typerende aanbieders op internet
- domeinenstructuur en url-opbouw
- verschillende soorten zoekstrategieën en -doelen
- verschil/toepassing gidsen versus zoekmachines
- werking en geavanceerde opties van zoekmachines
- betekenis van recall en precision
- betekenis diepe/onzichtbare web

# zo niet:

---

zie materiaal van oude cursus

"Internet Zoeken voor gevorderden"

- Sheets van Eric Sieverts:

<http://webzoek.googlepages.com/begin.htm>

- Startpagina van Jeroen Bosman:

<http://www.stgo.nl/kort/izg-start.htm>

of nieuwe cursus

"Internet als professionele informatiebron"

# soorten informatiebronnen

---

- naar aard van de informatie
  - wetenschappelijke informatie
  - naslagwerken
  - nieuwsberichten
  - krantenartikelen
  - (digitale) boeken
  - ...
- naar “medium”
  - webpagina’s
  - pdf’s
  - images
  - videos
  - weblogs
  - rss-feeds
  - ...
- onderscheid tussen deze invalshoeken voor indelen niet altijd heel scherp te trekken
- ook onderscheid: web ↔ diepe web

# soorten informatiebronnen

<i>niet al zulke combinaties zijn zinvol/mogelijk</i>	web-pagina	pdf	image	video	weblog	rss-feed
wetenschap						
naslag						
nieuws						
kranten						
eBooks						
...						



# informatiebronnen en hun zoekingen (*de zoektools*)

---

- kun je alles (ook) met Google vinden?
- welke Googles (en Yahoo's) zijn er allemaal?  
(*image-, blog-, video-, news-, book-, scholar-, groups-search*)
- kun je met die Googles alles vinden?
- welke alternatieven zijn er voor die Googles?  
(*zie bijv. intelways <http://www.intelways.com/> )*)

# 1. boeken & (wetenschappelijke) artikelen

---

- boeken
  - [Google Books](#)
  - [Amazon](#) (*ook reviews, inhoud, boek-boek citaties*)
  - [Worldcat](#) (*catalogus van 70.000 bibliotheken met postcode-functie*)
  - [Librarything](#) (*catalogus van 33.000.000 boeken van 500.000 bezitters*)
  - enz.
- artikelen (e.d.)
- enz.

# 1. boeken & (wetenschappelijke) artikelen (2)

---

- boeken
- artikelen (e.d.)
  - [Google Scholar](#) (artikelen, rapporten, proefschriften, ...)
  - [Scirus](#) (artikelen uit ElsevierScience, uit databases, + webpagina's, ...)
  - [OAster](#) / [Scientific Commons](#) (uit academische repositories)
  - [Pubmed](#) en dergelijke databases
  - [FindArticles](#), [Magportal](#) (ook -Engelstalige- publiekstijdschriften)
  - [ArticleFinder](#) (vooral uit wetenschappelijke tijdschriften)
  - enz.
- enz.

# boeken: Google Books

---



- Van kaft tot kaft gescand
- Sterk voor ontdekken van niet-hoofdinhoud van boeken
- Vaak beperkt doorbladerbaar  
*(no preview / snippet view / limited preview / full preview)*
- Via uitgevers en via grote bibliotheken
- Plaatsnamen uit boek >> Google-Maps
- Overlap met Amazon 'search inside'
- Probleem met copyrighted materiaal uit grote bibliotheken
- Vergelijking met Amazon: meer oud, minder recent
- Ook 'My Library' optie
- NL-boeken niet alleen uit Gent, ook uit US/UK

# wetenschap: Google Scholar



- > 100 miljoen wetenschappelijke publicaties
- concurrent van Web of Science, Scopus, Findarticles, Picarta, Scirus, OAlster, DOAJ, Infotrieve, Ingenta
- gericht geïndexeerd in vele bronnen (uitgevers, abstract-databases, universiteiten)
- met citaties!
- citaties mede gebruikt voor ranking
- advanced search beperkt, nog veel fouten
- van zoekresultaat naar volledige tekst steeds vaker een probleem (*niet aanwezig, alleen tegen betaling*)
- koppeling aan bibliotheekbezit, Google Scholar library programme
- geen info over bronnen, updates

Maak nu de opdrachten van onderdeel 1

## 2. feiten & naslag

---

- encyclopedieën e.d.
  - [wikipedia](#)
  - overzicht in [Yahoo Directory](#)
  - overzicht in [Open Directory](#)
  - overzicht in [encyclopedie.pagina.nl](#)
  - [internet movie database](#)
- vragen & antwoorden
  - [Al@din](#)
  - [Yahoo-answers](#)
  - [Live-QnA](#)
  - FAQs: [internet FAQ consortium](#)
- woordenboeken e.d.
- *enz.*

# wikipedia



WIKIPEDIA  
*The Free Encyclopedia*

- in [>250 talen](#)
- levert “wisdom of the crowds” altijd “wisdom”?
- goed voor “feitelijke” onderwerpen
- veel detailonderwerpen (*10 miljoen lemma's, 500.000 NL*)
- soms nuttig om relevante eigen publicaties / boeken uit eigen collectie als referenties aan lemma toe te voegen

## 2. feiten & naslag (2)

---

- encyclopedieën e.d.
- vragen & antwoorden
- woordenboeken e.d.
  - [answers.com](http://answers.com) (*voert een metasearch uit*)
  - [Roget thesaurus](#)
  - [Acronymfinder](#)
  - [Bartleby](#)
  - overzicht in [Open Directory](#)
  - overzicht in [Open Directory Nederlands](#)
- *enz.*

Maak nu de opdrachten van onderdeel 2

## 3. nieuws, kranten, tijdschriften

---

- [Google news](#)
- [Yahoo News](#)
- [BBC](#) , [CNN](#) *e.d.*
- [nu.nl](#)
- € [LexisNexis](#)
- krantensites (*overzicht: [startnederland.nl](#)*)
- tijdschriftensites (*overzicht: [startnederland.nl](#)*)
- [Newslink-magazines](#)
- ....

*[weblogs: zie onderdeel 4]*

Engelstalig nieuws uit 4500 bronnen

+ aparte versies in andere talen dan Engels:

- chinees (1000 bronnen)
- duits (700 bronnen)
- frans (500 bronnen)
- hebreuws (100 bronnen)
- italiaans (250 bronnen)
- japans (600 bronnen)
- koreaans (550 bronnen)
- nederlands (>400 bronnen)
- portugees (200 bronnen)
- spaans (700 bronnen)
- *enz.*



Maak nu de opdrachten van onderdeel 3

## 4. blogs & rss-feeds

---

- [Google blogsearch](#)
- [technorati](#)
- [icerocket](#)
- [bloglines](#)
- [blogpulse](#)
- [ask blogs&feeds](#)
- [exalead](#) >> *achteraf inperken op blogs*
  
- denk aan verschil tussen ingang op individuele berichten en ingang op blogs of feeds als geheel
- filtermogelijkheid op “autoriteit” ?=? belang, kwaliteit, ....

Maak nu de opdrachten van onderdeel 4

## 5. multimedia

---

- images
  - [Google-image](#) (*bevat ook foto's uit Picasa*)
  - [Yahoo-image](#) (*ook veel foto's uit Flickr*)
  - [Ask-image](#)
  - [Exalead-image](#)

*(veel verschillen in zoekresultaten door verschillen in “tekstzoekvelden”)*

  - [Flickr](#) (*zoekt vooral op tags; ondersteunt “Creative Commons”*)
  - [Tiltomo](#), [Retrievr](#) (*voorbeeldzoeken op Flickr*)
  - ...
- video
- podcasts (audio & video)

## 5. multimedia (2)

---

- images
- video
  - [Google-video](#) (*ook video's uit YouTube; zoekt alleen in metadata*)
  - [Gaudi](#) (*heel klein deel uit Google-video met spraakherkenning*)
  - [YouTube](#) (*groei: 13 uur/minuut*)
  - [EveryZing](#) (*door spraakherkenning full-text time-frame toegang; bevat ook veel materiaal uit YouTube*)
  - [Blinkx](#) (*26 miljoen uur video, spraakherkenning?*)
  - [Truveo](#)
  - [VideoSurf](#)
  - ...
- podcasts, audio
  - [EveryZing](#) (*door spraakherkenning full-text time-frame toegang*)
  - [Podscope](#) (*spraakherkenning*)
  - [Yahoo-audio](#) (*vooral muziek; geen spraakherkenning*)

Maak nu de opdrachten van onderdeel 5

## 6. archief & oud nieuws

---

- [web archive](#)  
*(oude versies van websites, terug tot 1996, ingang via -oude- url)*
- [Google news archive](#) *(historische krantenarchieven – in opbouw)*
- historische nederlandse kranten *(versnipperd aanbod)*
  - [collectie KB](#)
  - [overzicht](#) *(alleen gedigitaliseerd materiaal)*
  - [landelijke krantendatabank](#) *(ook papieren collecties)*
- € [LexisNexis](#)

Maak nu de opdrachten van onderdeel 6

## 7. gespecialiseerde versus algemene zoekingen

---

gespecialiseerde:

selectief & vaak “dieper” dan algemene zoekmachine

- interne zoekfunctie van individuele site
- kant en klare onderwerpsgerichte zoekingen
  - [Scirus](#) (wetenschap algemeen)
  - [Findlaw](#) (rechten USA)
  - [WebMD](#) / [Medscape](#) (geneeskunde)
  - [Google-finance](#) (financiële gegevens + nieuws)
  - .....
- homemade met:
  - [Google CSE](#), [Rollyo](#) of [Gigablast](#)  
(zie: <http://www.searchengineshowdown.com/cse/> )  
*zelf selecteren welke sites (url's) in zoekmachine moeten worden meegenomen*

## toegang tot het diepe web

- “handmatig” (elke database afzonderlijk doorzoeken)
  - overzicht "[Complete planet](#)" bij BrightPlanet
- “halfautomatisch” (doorzocht via metasearch / desktop search engine)
  - [turbo10](#) (*gespecialiseerde metasearch*)
  - [goshme](#) (*geavanceerde metasearch*) [**werkt tijdelijk (?) niet**]
  - [copernic](#) (*metasearch vanaf eigen PC*)
- “automatisch” (gegevens uit bepaalde database/bron zitten ook al -allemaal?- in algemene zoekmachines)  
*bijv. catalogus UB Wageningen*

Maak nu de opdrachten van onderdeel 7

## 8. vergelijken en beoordelen

---

vergelijken en beoordelen van (zoekingen van) bronnen

- wat voor soort materiaal zit er in
  - primaire (c.q. full-text) vs. secundaire informatie
  - naslag / nieuws / onderzoekresultaten / statistieken / websites / artikelen / boeken / rapporten / ...
- welk onderwerpsgebied
- welke doelgroep (*niveau*)
- hoeveel zit erin (*aantal items, dekkingsgraad, ...*)  
*vaak moeilijk te bepalen door verschillen in functionaliteit en onbetrouwbare aantallen voor zoekresultaten uit zoekmachines*
- hoe goed werkt het zoekstelsel  
*(functionaliteit, gebruiksgemak, selectiviteit, ...)*
- wat is de kwaliteit van de content  
*zie bijvoorbeeld [JISC Academic Database Assessment Tool](#)*

# beoordelingscriteria voor websites

---

- Goede opmaak
- Aanduiding maker/auteur (“about us”)
- Aanduiding postadres, telefoonnummer
- Aanduiding doel/doelgroep
- Geen reclame en pop-ups
- Heldere navigatie
- Interne zoekfunctie
- Voldoende snelheid server
- Backlinks door gezaghebbende organisaties
- Up to date?
- Zinnige datering inhoud
- Geen grof taalgebruik
- Geen kinderlijk taalgebruik
- Geen storende taalfouten

Zelfs als alles in orde lijkt, bij gevoelige onderwerpen toch nog uitkijken

Lookup: <http://www.allwhois.com> of  
<http://centralops.net/co/DomainDossier.aspx>

# bepaling van onze 20 beste brede “bronnen”: proefondervindelijk en met overleg in subgroepjes uit 40 brede bronnen de tien beste kiezen

---

- Al@din
- Amazon
- Answers.com
- Bartleby
- BBC
- Bibliotheek.nl
- Bloglines
- Britannica
- CBS-Statline
- EVD
- EveryZing
- Findarticles
- Geheugen van Nederland
- Google Groups
- Google Books
- Google Nieuws/News
- Google Scholar
- IMDB
- Infomine
- Infoplease
- Infotrieve Articlefinder
- Internet Public Library
- KB
- Kennisnet
- Muziekweb
- Nu.nl
- OAlster
- Omroep.nl
- Open Directory
- Overheid.nl
- Pubmed
- Startnederland
- Swets
- Teletekst
- VNG
- Wegener/GPD regionale dagbladen
- Wettenbank
- Wikipedia
- Worldcat
- Yahoo gids

## 9. bijblijven met bronnen

---

- Weblogs:
  - [Resourceshelf](#) (Gary Price)
  - [SearchengineLand](#) (Sullivan, Sherman)
  - [InternetNews/WebSearchGuide](#) (Gwen Harris)
  - [search engine showdown](#) (Greg Notess)
  - [IP Weblog](#) (Marten Hofstede)
- [Pandia](#)
- Kritische verhalen: [Digital Reference Shelf](#) (Jacso)

# bijblijven met behulp van RSS

---

- Rich Site Summary / Really Simple Syndication
- voor sites met (on)regelmatig nieuwe actuele inhoud, bijvoorbeeld nieuwsbrieven, weblogs (+ ook sommige databases)
- lezen, beheren als ware het email
- maakt gebruik van xml-structuur
- vereist software (reader), bijv. [FeedReader](#) (gratis) of [Feeddemon](#), of online lezen bijv. bij [Bloglines](#), [Google Reader](#) of [Netvibes](#)
- toevoegen van een 'feed': klik bij pagina op oranje (soms blauwe) rss- of xml-logo en voeg url daarvan aan reader toe
- drie nauwverwante formats: rss/atom/rdf (*meeste readers multiformat*)
- zoeken van berichten: Google Blogsearch, Technorati, IceRocket, Ask-blogs&feeds, Blogpulse [*zie eerder*]

