



retrieval en ontsluiting

taaltechnologische aanpak voor zoekproblemen

of eigenlijk:
"vind"-problemen

eric sieverts

informatiestructuren

najaar 2018

problemen met “vinden”

*in variabele mate in uiteenlopende soorten systemen:
catalogi, bibliografische databases, full-text bestanden,
enterprise search, het web, record management,*

- onvoldoende recall (*vangst*)
 - met zoekvraag mis je te veel relevante informatie
(je vindt te veel niet)
- onvoldoende precisie
 - zoekvraag levert (te) veel niet-relevante informatie
(je vindt te veel -ten onrechte- wel)

oorzaken voor lage recall (*de recall-killers*)

inherent aan free-text zoeken in documenten:

- variatie in spelling en woordvorm van gebruikte woorden
- tekst bevat synoniemen, impliciete aanduidingen, woorden in andere taal, ...
- tekst bevat variëteit aan mogelijke (specifiekere) woorden voor generiek begrip waarop is gezocht
- term-armoede van documenten (*catalogi, niet-tekst docs, ...*)

schuld van de zoeker:

- verkeerde zoekterm (*ongebruikelijk, foute spelling*)
- te weinig varianten met **OR** gecombineerd
- te veel zoek-elementen met **AND** gecombineerd

oorzaken voor lage precisie (*de precisie-killers*)

inherent aan free-text zoeken in documenten :

- in gevonden tekst geen of onbedoeld verband tussen met **AND** gecombineerde termen (*false coordination*)
- zoekwoord komt in tekst in andere betekenis voor (*homografen, acroniemen, andere taal*)
- zoekwoord komt voor in andere context dan bedoeld
- term-rijkdom van full-text documenten (*laag term-gewicht*)

schuld van de zoeker :

- verkeerde zoekterm (*te algemeen, andere betekenis*)
- te weinig zoek-elementen met **AND** gecombineerd

klassieke oplossing :

gebruik van: thesaurus, classificatie / taxonomie

waarom kan dit een oplossing bieden?

- door browsen/navigeren kun je bij goede term/categorie uitkomen
- formaliseert betekenis (*homograaf-probleem*)
- kan verwijzen naar gebruikelijke(r) term (*vervangend zoeken*)
- kan term-rijkdom uniformeren
(*dus gelijk term-gewicht; o.a. door diepte-indexering*)
- legt inhoudelijke relaties tussen onderwerpen/termen (*generiek zoeken*)
- kan verband leggen tussen facetten van onderwerp (*precoördinatie*)

nadelen van klassieke oplossing

- gebrek aan flexibiliteit
(schrik van de vak-specialist)
- gebruik van kunstmatige informatietaal
(schrik van de ergonoom)
- duur omdat mensen termen moeten toekennen
(schrik van de manager)

(taal)technologische alternatieven

- truncatie, wordstemming, fuzzy search, samenstellingen
- suggereren van betere zoektermen
- semantische kennis toevoegen
- best-match zoeken met relevantie-ordering
- zoekresultaat clusteren in "domeinen/contexten"
- suggereren van aanvullende zoektermen
- terugkoppeling van zoekersoordeel

+ *nieuwe semantische technieken*



Web Images Maps Shopping More ▾ Search tools

About 3,720,000 results (0.46 seconds)

Veilig ondergronds | **Veiligheid** | Vialis

[www.vialis.nl/.../ibp.jsp?...Veiligheid...](#) - Translate this page

De Tunnelwet schrijft onder meer extra **veiligheidsvoorzieningen** en een permanente bewaking van **tunnels** voor. De 1.600 meter **lange** Hubertustunnel is ...

[Afdeling 6.9 Aanvullende regels tunnelveiligheid nieuwbouw en ...](#)

[vrom.bouwbesluit.com/Inhoud/.../afd6-9](#) - Translate this page

Dit artikel is ter implementatie van de richtlijn **tunnelveiligheid**. Omdat in Nederland dergelijke **lange tunnels** altijd van mechanische ventilatie voorzien zijn kan ...

Profiel: **Tunnels** - Nrc

[www.nrc.nl/W2/.../Tunnels/veiligheid.html](#) - Translate this page

Nederlandse Tunnels **Veilig**. Tunnelongevallen leiden meestal tot alle materiële schade. Bij brand: windkracht zeven. Echt **lange autotunnels** heeft Nederland ...

[NIFV - De veiligheid in wegtunnels een hele beleving](#)

[www.nifv.nl/web/show/id.../print=true](#) - Translate this page

Absolute **veiligheid** is niet te bereiken, maar alles bepalen op basis van ... Klein in vergelijking met de twintig tot dertig kilometer **lange tunnels** in Noorwegen of ...

[Extern onderzoek veiligheid Schipholtunnel | Omroep West](#)

[www.omroepwest.nl/.../extern-onderzoek-v...](#) - Translate this page

Nov 17, 2012 – ... onderzoek zich op de **veiligheidsprocedures** het treinmaterieel, logistieke processen en de situatie in de bijna 6 kilometer **lange tunnel** zelf.

[\[PDF\] De veiligheid in wegtunnels een hele beleving - Infopunt Veiligheid](#)

[www.infopuntveiligheid.nl/.../...](#) - Translate this page

File Format: PDF/Adobe Acrobat - Quick View

Channeltunnel is de aandacht voor de **veiligheid** van tunnels, gemeten aan de hand ... Klein in vergelijking met de twintig tot dertig kilometer **lange tunnels** in ...

Tunnelveiligheid - Rijkswaterstaat

[www.rijkswaterstaat.nl > Wegen > Veiligheid](#) - Translate this page

Voor **tunnelveiligheid** is daarom extra aandacht nodig. Een verlichte ... Nederland krijgt er dan ook steeds meer en steeds **langere tunnels** bij. Ook het verkeer ...

samenstellingen en woordvarianten

- Google breidt zoekactie automatisch daarmee uit

truncatie / stemming / fuzzy / samenstellingen

- trunceren computer* → computeronderwijs
[meestal actie en keuze van zoeker nodig]
- word- computer → computing, computation,
stemming computed, computers
communism → community, communication ??
sieverts → sievert ?? → betere recall
[vaak automatisch; uit te schakelen ?]
- fuzzy duivendak → duijvendak, duyvendak
serajevo → sarajevo, serajewo
gorbachov → gorbachev, gorbatjov, ...
kok → kop, kak ... ??
[ja/nee-keuze of mate van fuzziness aangeven]
- samenstellingen >>

compenseert variatie in woordvorm & spelling

maar pas op voor
ongewenste effecten !!

termen verbeteren

computer suggereert iets afwijkende term

- door fuzzy-search in woordenboek
- door fuzzy search in eigen index >> term die (veel) meer oplevert
- door generieke kennis van vaak gemaakte tikfouten of herkeningsfouten bij OCR
- door kennis van vaker gestelde zoekvragen

bijv.: Google ("showing results for ...")



Search Library

[Advanced Search](#)

Keep search refinements

New search

Search Results: Your search for **agralunocytosis** returned **0** results

Did you mean: [agranulocytosis](#)

Suggestions:

See if there are [results beyond your library's collection](#)

Go [back](#) to your previous search, and try different filters.

Make sure all words are spelled correctly.

Try different keywords.

Try more general keywords.

Try fewer keywords.



zoekmahcinegoeroe



Search

Instant is on ▾

About 23,900 results (0.20 seconds)

[Advanced search](#)

Everything

Images

Videos

News

Shopping

More

[Show search tools](#)

Showing results for **zoekmachine goeroe**. Search instead for **zoekmahcinegoeroe**

Zoekmachine Goeroe; Zoekmachine Marketing en Zoekmachine ... ☆ 🔍

- [[Translate this page](#)]

Zoekmachine Goeroe helpt u bij het analyseren van uw website en biedt u de mogelijkheid om uw website zoekmachine vriendelijk te krijgen.

www.zoekmachinegoeroe.com/ - [Cached](#)

SEM, SEO en SEA - **Zoekmachine Goeroe**; Zoekmachine Marketing en ... ☆ 🔍

- [[Translate this page](#)]

Zoekmachine Goeroe helpt u bij het analyseren van uw website en biedt u de ...

www.zoekmachinegoeroe.com/index.php?zoekmachine=seo-sea - [Cached](#)

Links - **Zoekmachine Goeroe**; Zoekmachine Marketing en Zoekmachine ... ☆ 🔍

- [[Translate this page](#)]

Zoekmachine Goeroe helpt u bij het analyseren van uw website en biedt u de ...

zoekmachinegoeroe.com/index.php?zoekmachine=links - [Cached](#)

PHP - **Zoekmachine Goeroe**; Zoekmachine Marketing en Zoekmachine ... ☆ 🔍

- [[Translate this page](#)]

19 aug 2010 ... **Zoekmachine Goeroe** helpt u bij het analyseren van uw website ...

zoekmachinegoeroe.com/index.php?zoekmachine=php - [Cached](#)

[+](#) [Show more results from zoekmachinegoeroe.com](#)

Zoekmachine optimalisatie en Webdesign Stadskanaal ☆ 🔍 - [[Translate this page](#)]

SEO **goeroe** en **Zoekmachine** optimalisatie specialist Ron Wubs Webdesign Meer dan 12 jaar ICT ervaring. W3C goedgekeurde websites. ...

www.ronwubs.nl/ - [Cached](#) - [Similar](#)

De **zoekmachine** marketing **goeroe** voor alle bedrijven en websites ☆ 🔍

- [[Translate this page](#)]

Rolf is de **zoekmachine** marketing **goeroe** voor bedrijven die zich online willen profileren op de nationale en internationale markt, email hem met uw verzoek.

www.nederland-marketing.com/ - [Cached](#)

Nieuws: SEO Congres: Keesjan Deelstra, de **zoekmachine goeroe** ... ☆ 🔍

- [[Translate this page](#)]



zoekmachinegeoroe



Search

Instant is on ▾

About 62,900 results (0.16 seconds)

[Advanced search](#)

Everything

Images

Videos

News

Shopping

More

All results

[Wonder wheel](#)

[Timeline](#)

[More search tools](#)

Showing results for **zoekmachine george**. Search instead for **zoekmachinegeoroe**

[Barack Obama wordt 'miserable failure' als zoekmachines er niets ...](#) ☆ 🔍

- [[Translate this page](#)]

Als **zoekmachines George** W. Bush laten zien bij de zoekterm miserable failure, hoe kan het dan dat Obama hetzelfde lot beschoren is? Google zegt inmiddels de ...
[www.traffic-builders.com/.../barack-obama-wordt-miserable-failure-als-zoekmachines-er-niets-aan-doen.html](#) - [Cached](#) - [Similar](#)

[Truveo video zoekmachine vernieuwd - George Michael Forums](#) ☆ 🔍

1 post - 1 author - Last post: 16 Jun 2008

Truveo video **zoekmachine** vernieuwd Music | Celeb RSS News.
[www.georgemichaelforums.com > ... > Music | Celeb RSS News](#) - [Cached](#)

[Booleaanse algebra van George Boole | Wetenschap: Wiskunde](#) ☆ 🔍

- [[Translate this page](#)]

Veel gebruikers van **zoekmachines** komen bij hun antwoorden door zijn theorie, al zijn zij zich hier zelf niet van bewust. De **zoekmachine**, met dank aan **George ...**
[wetenschap.infonu.nl/.../38235-booleaanse-algebra-van-george-boole.html](#) - [Cached](#) - [Similar](#)

[Open - Nederlandse zoekmachine - DutchJoomla! Forum](#) ☆ 🔍 - [[Translate this page](#)]

4 posts - 3 authors - Last post: 18 Oct 2009

Open - Nederlandse **zoekmachine** Joomla Extensies. ... Standaard Re: Nederlandse **zoekmachine ...** Stuur een privébericht naar **George V. ...**
[forum.dutchjoomla.org > Joomla! 1.5.x > Joomla Extensies](#) - [Cached](#)

[NHV - Zoekmachine - Sporthal - NHV-DISTRICT-NOORDHOLLANDNOORD ...](#) ☆ 🔍

- [[Translate this page](#)]

9688, 10:00, EJ, 04, **George** St. E1, -, Sew Kiddy World E2 ... 8067, 13:30, DBJ, 3A, **George** St. B1, -, Valken Alpha Tours B3 ...
[www.noord.nhv.nl/zoekmachine/sporthal/NHV.../ZQ129MT?...17...](#) - [Cached](#)

[NHV - Zoekmachine - Sporthal - NHV-DISTRICT-NOORDHOLLANDNOORD ...](#) ☆ 🔍

- [[Translate this page](#)]

8677, 12:45, DCJ, 10, **George** St. C2, -, Risdam Zebra's C1 ...
[www.nhv.nl/zoekmachine/sporthal/NHV.../ZQ129MT](#) - [Cached](#) - [Similar](#)

[Web](#)[Images](#)[News](#)[Videos](#)[Shopping](#)[More ▾](#)[Search tools](#)

About 29,500 results (0.76 seconds)

Showing results for **zoekmachine goeroe**

No results found for **zoekmachinegoeroe**

*maar Google heeft
intussen bijgeleerd*

✓ [Zoekmachine Goeroe](#)

www.zoekmachine-goeroe.nl/ ▾ [Translate this page](#)

Index of / .ftpquota · cgi-bin/

✓ [SEOguru | SEO trainingen, onderzoek en advies Amsterdam](#)

www.seoguru.nl/ ▾ [Translate this page](#)

Kennis-, training- en adviescentrum voor moderne, betrouwbare **zoekmachine** optimalisatie (SEO) voor Google.

✓ [Zoekmachinegoeroe: Zoekmachine vriendelijke websites](#)

www.zoekmachinegoeroe.com/ ▾ [Translate this page](#)

Zoekmachinegoeroe de echte nr 1 van Nederland. Lees meer over zoekmachine vriendelijke websites.

✓ [zoekmachine-goeroe.nl - WooRank.com](#)

<https://www.woorank.com/en/www/zoekmachine-goeroe.nl>

We are collecting data and creating a prioritized Marketing Checklist for **zoekmachine-goeroe.nl**. Please allow up to 24 hours for all of your data to be processed.

✓ [Zoekmachine Goeroe | Facebook](#)

<https://www.facebook.com/zoekmachine.goeroe> ▾ [Translate this page](#)

Zoekmachine Goeroe is on Facebook. Join Facebook to connect with **Zoekmachine**

termen verbeteren

computer suggereert iets afwijkende term

- door fuzzy-search in woordenboek
- door fuzzy search in eigen index op een term die (veel) meer oplevert
- door generieke kennis van vaak gemaakte tikfouten of herkeningsfouten bij OCR
- door kennis van vaker gestelde zoekvragen

bijv.: Google ("showing results for ...")

semantische kennis

semantisch netwerk

bevat het hele woordenboek – geen voorkeurstermen

- betekenissen van woorden worden vastgelegd
(ook bij meer betekenissen)
- er worden betekenisrelaties gelegd tussen inhoudelijk verwante woorden *(in één of meer talen)*
- omgeving van woord in het netwerk maakt het mogelijk betekenissen te onderscheiden (in document en in query)
- omgeving van woord in het netwerk kan termen leveren om query mee te expanderen *(woorden binnen bepaalde "semantische afstand" van zoekwoord)*

voorbeeld: [Wordnet](#) *(voor het Engels)*

visualisatie van "wordnet"

The screenshot shows a software interface for a Visual Thesaurus. At the top, there is a search bar with the word "profession" and a "LOOK IT UP" button. To the right are buttons for "SEARCH: EN", "EDIT", "PRINT", and "HELP". Below the search bar, the word "job" is entered in the search field, and the interface displays a network of related words.

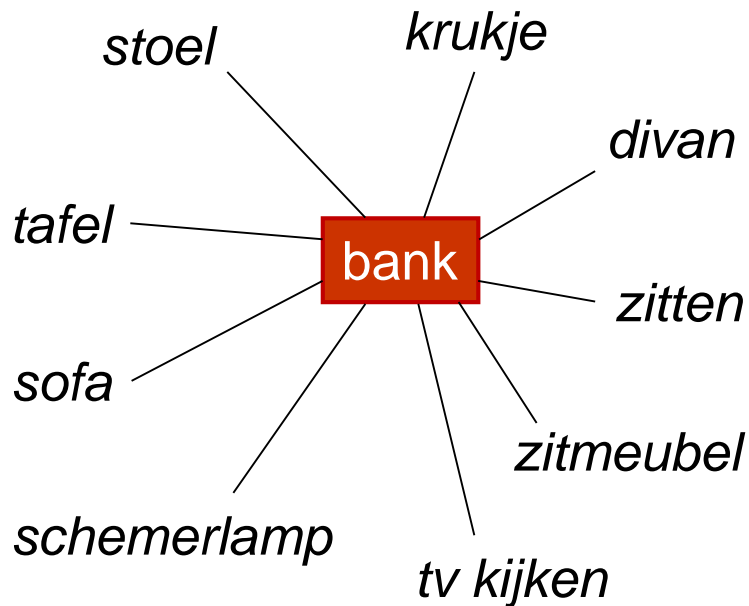
The network diagram has "job" at the center. It is connected to various words, some of which are further connected to other words, forming a web. The words are color-coded: red for nouns, green for verbs, and purple for adverbs. A callout box points to the word "cheat" with the definition: "a crime (especially a robbery) 'the gang pulled off a bank job in St. Louis'".

On the right side, there is a list of related words categorized by part of speech:

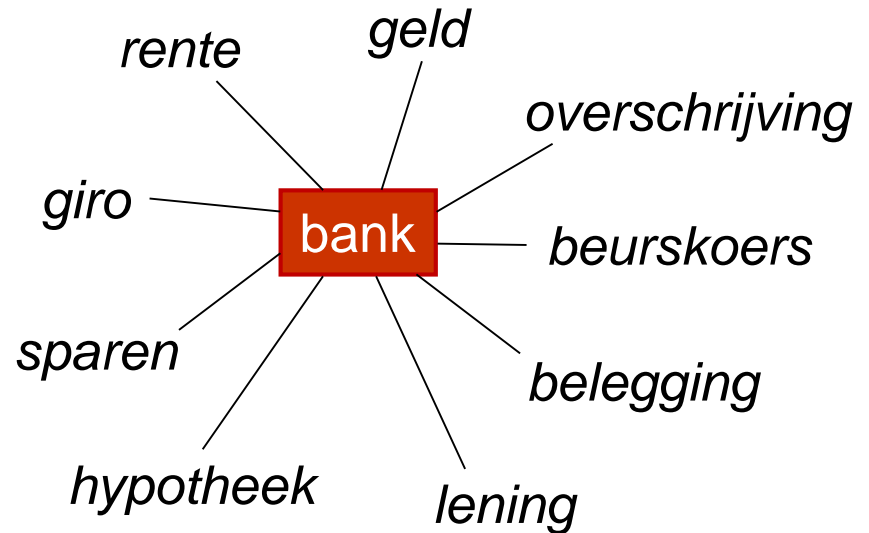
- NOUNS** (ON):
 - the principal activity in your life that you do to earn money
 - a specific piece of work required to be done as a duty or for a specific fee
 - the performance of a piece of work
 - the responsibility to do something
 - a workplace; as in the
- ADJECTIVES** (ON)
- VERBS** (ON):
 - profit privately from public office and official business
 - arranged for contracted work to be done by others
 - work occasionally
 - invest at a risk
 - be employed
 - engage in deceitful behavior; practice trickery or fraud
- ADVERBS** (ON)

At the bottom of the screen, there is a footer: "VISUAL THESAURUS (R) VERSION 3 | CREATED USING THINKMAP (R) | COPYRIGHT 2005 THINKMAP INC."

betekenissen onderscheiden



*"hij zat op de **bank** tv te kijken,
met een biertje voor zich op tafel"*



*"deze **bank** rekent een hoge rente
op leningen en hypotheek"*

*"onder het schijnsel van de schemerlamp zat hij op de **bank** een
giro uit te schrijven om bij zijn **bank** de hypotheek af te lossen"*

synoniemen toevoegen

Articles

Case law

My library

Any time

Since 2016

Since 2015

Since 2012

Custom range...

Sort by relevance

Sort by date

 include patents include citations Create alert**Gender differences** in factors influencing alcohol use and **drinking** progression among **adolescents**MT Schulte, [D Ramo](#), SA Brown - Clinical psychology review, 2009 - Elsevier

While prevalence rates for alcohol use and related disorders **differ** widely between adult men and women, male and female **adolescents** do not exhibit the same disparity in alcohol consumption. Previous research and reviews do not address the emergence of ...

Cited by 150 [Related articles](#) [All 10 versions](#) [Cite](#) [Save](#)[\[HTML\] from nih.gov](#)
[Fulltext@UBUlink](#)Towards the assessment of **adolescent** problem **drinking**.[HR White](#), [EW Labouvie](#) - Journal of studies on alcohol, 1989 - jsad.com

Comparisons of participants and eligible nonparticipants on various demographic variables revealed an overrepresentation of higher parental ... In **comparison**, the strength of the relationship between use intensity and RAPI tends to fall between ... **Sex differences** and age trends ...

Cited by 1152 [Related articles](#) [All 10 versions](#) [Cite](#) [Save](#)[\[PDF\] from researchgate.net](#)
[Fulltext@UBUlink](#)A longitudinal study of antisocial behaviors in early **adolescence** as predictors of late **adolescent** substance use: **gender** and ethnic group **differences**.

M Windle - Journal of abnormal psychology, 1990 - psycnet.apa.org

... To minimize the Type I error rate due to multiple **comparisons**, the alpha level for statistical significance was set at .01. ... **Sex differences** in the antecedents of substance use among **adolescents**. ... Normal deviance: **Gender** similarities in **adolescent** subcultures. ...

Cited by 327 [Related articles](#) [All 11 versions](#) [Cite](#) [Save](#)[Fulltext@UBUlink](#)**Gender differences** and areas of common concern in the driving behaviors and attitudes of **adolescents**[N Harré](#), J Field, B Kirkwood - Journal of Safety Research, 1996 - Elsevier

... Risk for DWI: A new look at **gender differences** in **drinking** and driving influences, experiences, and attitudes among new **adolescent** drivers. ... **Sex differences** in driving performance. ... Sensation seeking in England and America: Cross cultural, age and **sex comparisons**.

Cited by 183 [Related articles](#) [All 8 versions](#) [Cite](#) [Save](#)[Fulltext@UBUlink](#)A **gender**-specific measure of binge **drinking** among college students.

H Wechsler, GW Dowdall... - American journal of ..., 1995 - ajph.aphapublications.org

... By contrast, for eight of the nine problems reported (the column labeled "Gender-Specific Definition"), women who typically drank four ... association between usual **drinking** and the occurrence of alcohol-related problems might be owing not to **sex** but to **differences** in body ...

Cited by 650 [Related articles](#) [All 13 versions](#) [Cite](#) [Save](#)[\[PDF\] from nih.gov](#)
[Fulltext@UBUlink](#)Motivations for alcohol use among **adolescents**: Development and validation of a four-factor model.

ML Cooper - Psychological assessment, 1994 - psycnet.apa.org

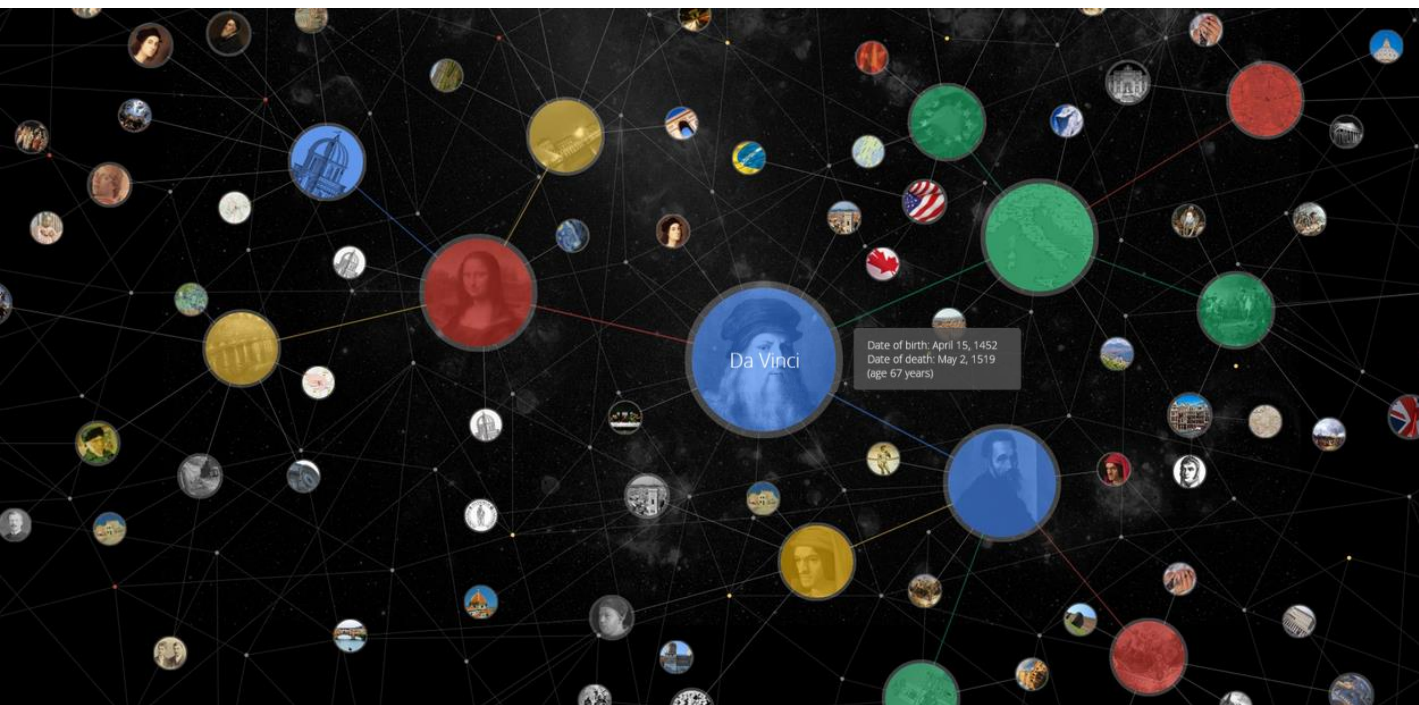
... Assuming adequate evidence of factor structure invariance, mean **differences** in **drinking** motives

[\[PDF\] from researchgate.net](#)
[Fulltext@UBUlink](#)

semantische kennis

Google "Knowledge Graph" (en Bing's "entity search")
nog geavanceerder netwerken met veel meer semantische
kennis, waarmee zoekacties (ongemerkt) verbeterd worden

- helaas voor gebruiker niet zichtbaar wat precies gebeurt



**Knowledge
Graph kent
>2 miljard
objecten met
>70 miljard
kenmerken
en relaties**

semantische kennis

in semantisch netwerk

- worden betekenissen van woorden vastgelegd (*het hele woordenboek*)
- worden betekenisrelaties gelegd tussen inhoudelijk verwante woorden (*in één of meer talen*)
- omgeving van woord in het netwerk maakt het mogelijk betekenissen te onderscheiden (in document en in query)
- omgeving van woord in het netwerk kan termen leveren om query mee te expanderen (*woorden binnen bepaalde "semantische afstand" van zoekwoord*)

verbetert precisie

verbetert recall

**maar semantisch netwerk
voor specialistisch domein
moet je zelf nog bouwen/aanvullen**

relevance ranking factoren

1. meer van de gevraagde termen in een document
2. gevraagde termen op belangrijke plek in document (titel, url, koppen, eerste paar regels,)
3. gevraagde termen komen in document herhaald voor
4. gevraagde termen staan in document dicht bij elkaar
5. termen staan in document in zelfde volgorde als in vraag »
6. zeldzame termen krijgen zwaarder gewicht dan algemene
7. hoeveel hyperlinks naar document verwijzen
8. hoe vaak een document / site wordt "bezoekt"
9. of document lijkt op wat je eerder bekeken hebt
10. ...

werkt goed op het web, maar ook op een intranet?

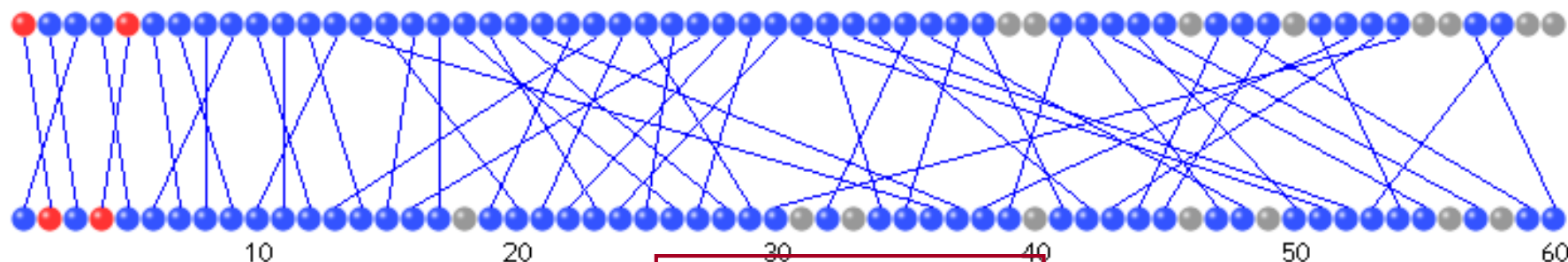
Visit: [Products](#) > [Thumbshots Images](#) > [Ranking](#)

Thumbshots Ranking compares your site ranking on

Search: versus

Highlight Site: (optional eg. mysite.com)

[YAHOO](#) : modern american composers



[YAHOO](#) : composers american modern

relevance ranking factoren

1. meer termen → meer concepten ge-AND
2. termen in titel/kop/begin → hoger term-gewicht
3. termen herhaald → hoger term-gewicht
4. termen dicht bij elkaar → juiste verband
5. termen in volgorde → juiste verband
6. zeldzame termen zwaarder → belang specifieke term
7. hyperlinks naar document → (kwaliteit) *[alleen als er links zijn]*
8. bezoek aan document → (kwaliteit)
9. persoonlijke voorkeur → juiste betekenis/context

relevance ranking factoren

1. meer van de gevraagde termen in een document
2. gevraagde termen op belangrijke plek in document (titel, url, koppen, eerste paar regels, ...)
3. gevraagde termen komen in document herhaald voor
4. gevraagde termen staan in document dicht bij elkaar
5. termen staan in document in zelfde volgorde als in vraag
6. zeldzame termen krijgen zwaarder gewicht dan algemene
7. hoeveel hyperlinks naar document verwijzen
8. hoe vaak een document / site wordt "bezoekt"
9. of document lijkt op wat je eerder bekeken hebt

allemaal gericht op hogere relevantie voor "de eerste tien", dus op precisie

automatisch clusteren/classificeren

- op grond van kennisregels
(en bestaande “taxonomie”)
in feite toepassing van automatische rule-based classificatie, waarbij klassen als verdelingscriterium dienen
- op grond van statistiek of patronen
 - [Carrot2](#), [Yippy](#),
 - [Autonomy](#)



clinton

Search

More options

Folders Circles FoamTree

- All Topics (100)
 - Secretary of State (9)
 - Clinton County (8)
 - Hillary Clinton's Emails (8)
 - Donald Trump (7)
 - President Bill Clinton (7)
 - Hathi Trust Digital Library (6)
 - Iowa (5)
 - North at Clinton (5)
 - William J. Clinton (5)
 - Day Ago (4)
 - more | show all

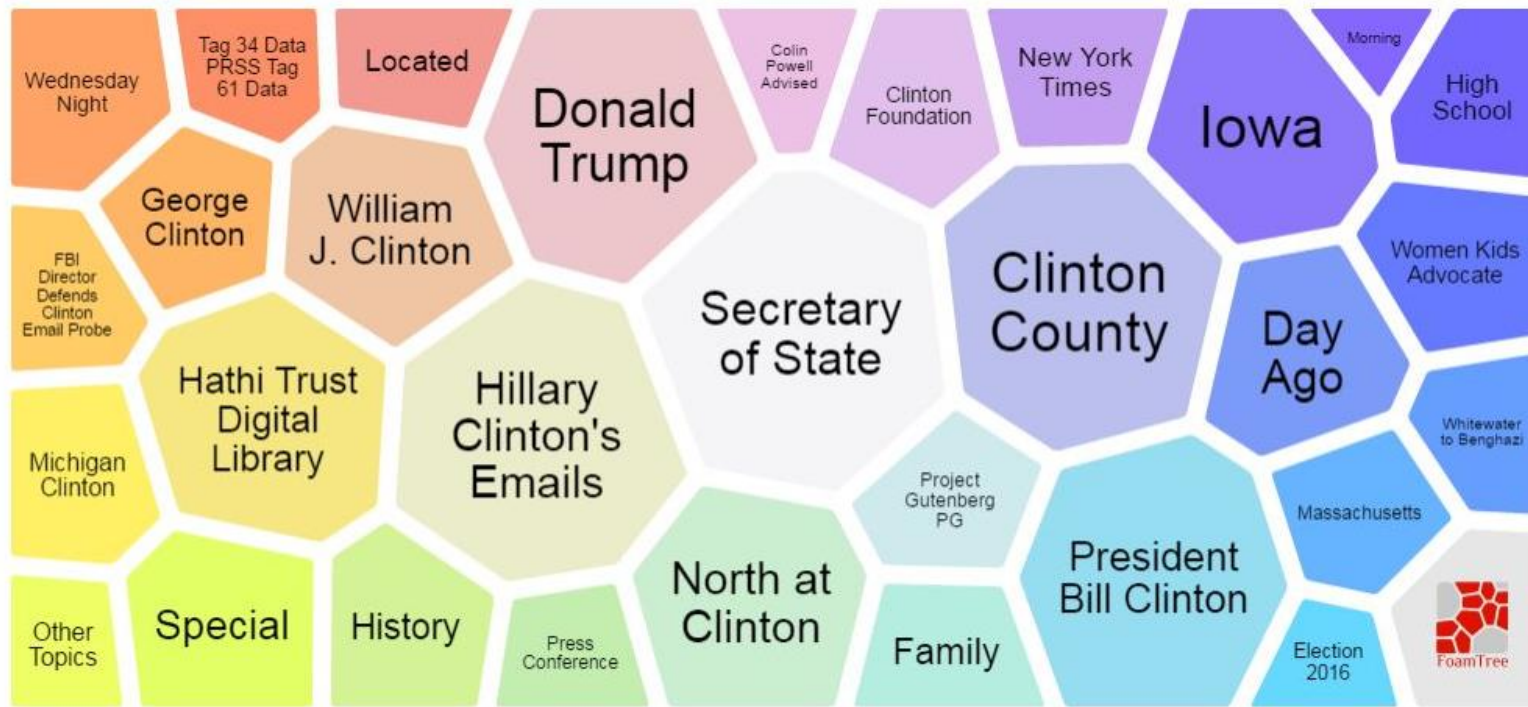
Top 100 results of about 100 for clinton

- [Hillary Clinton 2016 | Hillary for America](#)

The official website for Hillary **Clinton's** 2016 presidential campaign ...
<https://www.hillaryclinton.com/> [Ask, Bing, Goo, Google, Yahoo]
- [Bill Clinton - Wikipedia, the free encyclopedia](#)

Second term. Campaign · 1996 · 2nd inauguration · Operation Infinite Reach · Bombing of In H attended St. John's Catholic Elementary School, Ramble Elementary School, and Hot Springs High ...
https://en.wikipedia.org/wiki/Bill_Clinton [Ask, Goo, Google, Wikipedia]

Folders Circles FoamTree



Query: clinton -- Source: e



automatisch clusteren/classificeren

- op grond van kennisregels
(en bestaande “taxonomie”)

*in feite toepassing van regelgebaseerde classificatie, waarbij klassen als
verdelingscriterium dienen*

**kiezen van juiste betekenis of context
→ betere precisie**

- op grond van statistiek of patronen
 - Carrot2, Yippy, ...
 - Autonomy

**werkt niet gegarandeerd
altijd goed**

aanvullende termen extraheren

computer haalt karakteristieke (andere) woorden/begrippen uit eerste N documenten van zoekresultaat (*statistiek - $tf*idf$*)

- tf = term frequentie
computer turt van elk woord hoe vaak het in document voorkomt
- idf = inverse document frequentie
computer zoekt in zoekmachine-index op hoeveel documenten er zijn waarin dat woord voorkomt (df)
- computer deelt term-frequentie door document frequentie
(*"invers = delen door"*)
dat compenseert voor hoge term frequentie van algemeen voorkomende woorden

tf*idf

rekenvoorbeeld:

woord	tf	df	tf*idf
de	30	40.000	0,00075
in	15	30.000	0,00050
karakteristiek	2	40	0,050
zoekmachine	1	15	0,067

in praktijk vaak nog verfijning door in idf een logaritme te verwerken:

$$\text{idf} = \log \frac{N}{\text{df}} \quad \text{waar } N = \text{totaal aantal documenten}$$

aanvullende termen extraheren

die berekening levert rijtje termen,
gesorteerd op hun $tf*idf$ -waarde

- hoe hoger $tf*idf$ -waarde, hoe karakteristieker

- gebruiker kiest daaruit term om zoekactie in te perken
(soms ook termen uit te sluiten, of juist resultaat met
OR uit te breiden)

bijv.: Quintura (vroegere demo met frontend voor Yahoo)

Aquabrowser (vroeger bij bibliotheek.nl)



Documents found - 319000000

1. **Stone** - Wikipedia, the free encyclopedia

Dimension **stone**, **stone** fabricated to specific sizes or shapes. Calculus (medicine), a **stone** formed in the body, such as kidney **stones** or gallstones ...

<http://en.wikipedia.org/wiki/Stone>

2. **The Rolling Stones**

Official site of the legendary rock band The Rolling **Stones**. Includes news, tour information, music, photos, and historical archives.

<http://www.rollingstones.com/>

3. **Kidney stone** - Wikipedia, the free encyclopedia

The **stones** are solid concretions or calculi (crystal aggregations) formed in the kidneys ... The term bladder **stones** usually applies to urolithiasis of the ...

http://en.wikipedia.org/wiki/Kidney_stone

4. **Kidney stones**: Medlineplus Medical Encyclopedia

These substances can create small crystals that become **stones**. ... When this happens, the **stones** can block the flow of urine out of the kidneys. ...

<http://www.nlm.nih.gov/medlineplus/ency/article/000458.htm>

"kidney stones"

Web Images Video

Site search
 Search for kids

mayoclinic **treatment** department
 kosmix new urologychannel symptom medpedia
 fact natural health medical aol
 wikipedia calculus herb medlineplus
 definition herb emedicinehealth

Documents found - 13500000

1. **Kidney Stones** in Adults

Discusses the causes, symptoms, and diagnosis of **kidney stones**. Reviews prevention strategies and treatment options, as well as information on current research.

<http://kidney.niddk.nih.gov/kudiseases/pubs/stonesadults/i...>

2. urologychannel: **Kidney Stones**

Kidney stones are **stones** formed from chemicals in urine that crystallize and stick together. Some cause problems, and others we may never know they're there.

<http://www.urologychannel.com/kidneystones/index.shtml>

3. **Kidney stone** - Wikipedia, the free encyclopedia

Kidney stones (ureterolithiasis) result from **stones** or renal calculi ... **Kidney stones** typically leave the body by passage in the urine stream, and many **stones** ...

http://en.wikipedia.org/wiki/Kidney_stone

4. **kidney stones?** what if i have them, do i need surgery? how do ...

Kidney stones are **stone**-like lumps that can develop in one or both of the **kidneys**. **Kidney stones** are quite common and, in the UK, they will affect ...

<http://answers.yahoo.com/question/index?qid=20080319041514...>

"rolling stones"

Web Images Video

Site search
Search for kids

artistdirect item rock ticket poster
sheet music music new official
mp3 download video
album lyric song artist category
biography free

Documents found - 7810000

1. **The Rolling Stones**

Official site of the legendary rock band The **Rolling Stones**. Includes news, tour information, music, photos, and historical archives.

<http://www.rollingstones.com/>

2. **Rolling Stones**

<http://www.stones.com/index.htm>

3. **Latest Rolling Stones News | Rollingstones.com**

The **Rolling Stones'** "She Said Yeah," a track from their album December's Children (And Everybody's) that was recorded and released in 1965, serves ...

<http://www.rollingstones.com/news>

4. **The Rolling Stones on Myspace Music - Free Streaming Mp3s ...**

Download The **Rolling Stones** Rock // music singles, watch music videos, listen to free streaming mp3s, & read The **Rolling Stones's** blog. ...

<http://www.myspace.com/therollingstones>

5. **The Rolling Stones - AOL Music**

The **Rolling Stones's** AOL Music web site, featuring The **Rolling Stones** news, The **Rolling Stones** music videos, The



ook “faceted” zoeken, waarbij zoekresultaat al wordt opgedeeld aan de hand van daarin aanwezige "geformaliseerde metadata"



urgent. Het Land in Saertinge vraagt niet minder om een inpoldering dan de Lauwerszee. ...

Let op: alleen in te zien in de bibliotheek



16. Waarom moet je eerst zoveel pijn lijden door ween als je een baby krijgt?

... prikkels worden doorgegeven zijn verschillend bij de ontsluiting enerzijds en de uitdrijving anderzijds. Laten we eens kijken wat er in beide gevallen gebeurt. De pijn bij de

wolk van termen in Aquabrowser: die termen kunnen uit statistische analyse, woordenlijst, thesaurus, semantisch netwerk o.i.d. komen

Verfijnen

Alle Bronnen

- Landelijke kranten (7)
- Websites (4)
- Onderwijs (3)
- Boeken (2)
- Al@din (2)

Boeken

- Non-fictie (2)
- Volwassenen (2)

Kranten

- Het Financiële Dagblad (3)
- De Volkskrant (2)
- NRC Handelsblad (1)
- Het Parool (1)

Onderwijs

- Davindi (3)

Websites

- Davindi+ (3)
- Internet Gids (1)

aanvullende termen extraheren

computer haalt karakteristieke (andere) woorden/begrippen uit eerste N documenten van zoekresultaat ($statistiek = tf * idf$)

- gebruiker kiest daaruit termen om zoekactie in te beperken (soms ook termen uit te sluiten of juist resultaat met OR uit te breiden)

inperken op juiste betekenis of context
→ betere precisie

uitbreiden met meer "synoniemen"
→ verbetert recall

terugkoppeling

gebruiker klikt bij relevante hit op “*more like this*” / “*similar*”

- computer zoekt op grond van daarin aanwezige termen of patronen naar daarop lijkende documenten

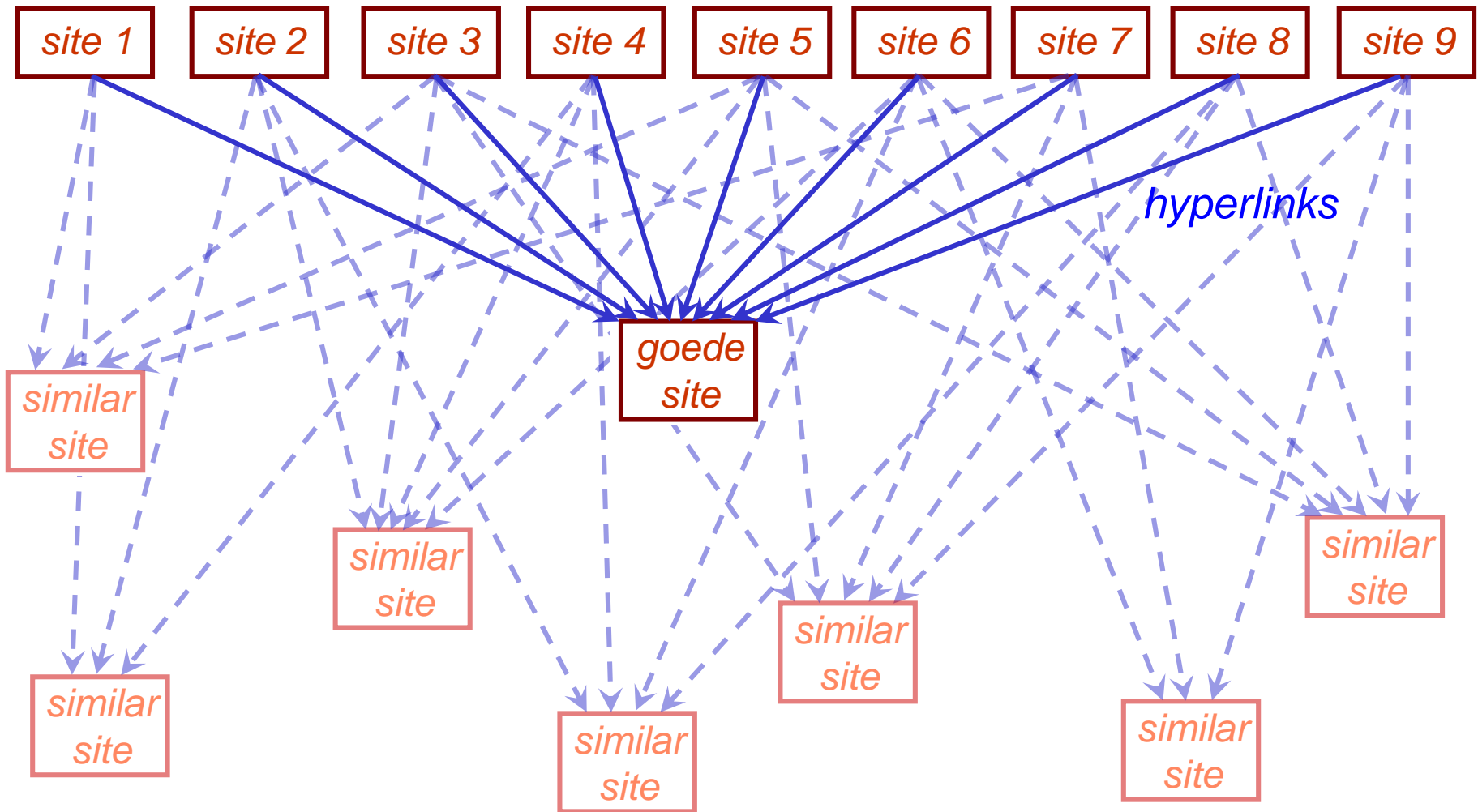
bijv.: *Google*
 Autonomy

gebruiker markeert relevante hits

- zoektermen of patronen die in die documenten voorkomen krijgen hoger gewicht
 - bij toekomstige zoekacties krijgen documenten met die termen of patronen een hogere berekende relevantie

bijv.: *Autonomy*
 en ook personal Google

gelijkenis-“patronen” bij Google



[Web](#)[Images](#)[Maps](#)[Shopping](#)[Books](#)[More ▾](#)[Search tools](#)

About 704,000 results (0.22 seconds)

[Max Havelaar - Wikipedia, the free encyclopedia](#)en.wikipedia.org/wiki/Max_Havelaar

Max Havelaar: Or the Coffee Auctions of the Dutch Trading Company (Dutch: **Max Havelaar**, of de koffi-veilingen der Nederlandsche Handel-Maatschappij) is a ...

[Max Havelaar France | Commerce équitable](#)www.maxhavelaarfrance.org/ ▾ [Translate this page](#)

Délivre un label garantissant aux consommateurs de café et de certains autres produits tropicaux que le circuit de commercialisation mis en place assure au ...

[Max Havelaar Switzerland](#)www.maxhavelaar.ch/ ▾ [Translate this page](#)

Newsroom. Aktuelle Meldungen. Das Neuste rund um **Max Havelaar** und den fairen Handel immer im Newsroom. **Max Havelaar** Neue Produkte ...

[Max Havelaar](#)www.maxhavelaar.com/ ▾

MaxHavelaar · Belgium · Denmark · France · Norway · The Netherlands · Swiss.

[Max Havelaar Foundation | Max Havelaar](#)www.maxhavelaar.nl/english ▾

You are on the website of the **Max Havelaar** Foundation. **Max Havelaar** is the independent campaigning and certification organization behind the Fairtrade Mark.

[Max Havelaar: Home](#)www.maxhavelaar.nl/ ▾ [Translate this page](#)

Keurmerk voor eerlijke handel. Maatschappelijk verantwoord ondernemen op basis van een duurzaam inkoopbeleid. Naast informatie over de organisatie ook ...

[Images for "max havelaar"](#) - Report images

See results about

[Max Havelaar](#)

Book by Multatuli
Max Havelaar: Or the Coffee Auctions of the ...

[Max Havelaar-Stiftung Schweiz](#)

The Max Havelaar Foundation is a ...

[Association Max Havelaar...](#)

Association Max Havelaar France is the French member of FLO ...

[Stichting Max Havelaar Company](#)

Stichting Max Havelaar is the Dutch member of ...

[Max Havelaar](#)

1976 Film
Max Havelaar is a 1976 Dutch drama film ...

**Web**

Images

Maps

Shopping

More ▾

Search tools



15 results (0.15 seconds)

[Max Havelaar: Or the Coffee Auctions of the Dutch Trading ...](#)[www.amazon.com](#) > Books > Literature & Fiction > Classics ▾

Max Havelaar: Or the Coffee Auctions of the Dutch Trading Company (Penguin Classics) [Multatuli, Roy Edwards, R. P. Meijer] on Amazon.com. *FREE* super ...

[Max Havelaar \(1976\) - IMDb](#)[www.imdb.com/title/tt0074880/](#) ▾

★★★★☆ Rating: 6.9/10 - 579 votes

Directed by Fons Rademakers. With Peter Faber, Sacha Bulhuis, Adendu Soesilaningrat, Maruli Sitompul. An idealistic Dutch colonial officer posted to ...

[Max Havelaar - 'Multatuli' - Complete Review](#)[www.complete-review.com/reviews/niederld/multatuli.htm](#) ▾

A review, and links to other information about and reviews of **Max Havelaar** by Multatuli.

[Max Havelaar Foundation | Max Havelaar](#)[www.maxhavelaar.nl/english](#) ▾

You are on the website of the **Max Havelaar** Foundation. **Max Havelaar** is the independent campaigning and certification organization behind the Fairtrade Mark.

[Max Havelaar Switzerland](#)[www.maxhavelaar.ch/](#) ▾ [Translate this page](#)

Newsroom. Aktuelle Meldungen. Das Neuste rund um **Max Havelaar** und den fairen Handel immer im Newsroom. **Max Havelaar** Neue Produkte ...

[Max Havelaar by Multatuli - Free Ebook - Project Gutenberg](#)[www.gutenberg.org](#) > 43,783 ebooks > 7 by Multatuli ▾

Free

Feb 1, 2004 - Free kindle book and epub digitized and proofread by Project Gutenberg.

[Max Havelaar by Multatuli \(Eduard Douwes Dekker\) | Iris on Books](#)[irisonbooks.com/.../max-havelaar-by-multatuli-eduard-douwes-dekker/](#) ▾

About 704,000 results (0.22 seconds)

[Max Havelaar - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Max_Havelaar ▾

Max Havelaar: Or the Coffee Auctions of the Dutch Trading Company (Dutch: **Max Havelaar**, of de koffi-veilingen der Nederlandsche Handel-Maatschappy) is a ...

[Background](#) - [Film version](#) - [See also](#) - [References](#)

[Max Havelaar France | Commerce équitable](#)

www.maxhavelaarfrance.org ▾ [Translate this page](#)

Délivre un label garantissant aux consommateurs de café et de certains autres produits tropicaux que le circuit de commerce équitable est en place assure au ...

[Max Havelaar Switzerland](#)

www.maxhavelaar.ch/ ▾ [Translate this page](#)

Newsroom. Aktuelle Meldungen. Das Neuste rund um **Max Havelaar** und den fairen Handel immer im Newsroom. **Max Havelaar** Neue Produkte ...

[Max Havelaar](#)

www.maxhavelaar.com/ ▾

MaxHavelaar · Belgium · Denmark · France · Norway · The Netherlands · Swiss.

[Max Havelaar Foundation | Max Havelaar](#)

www.maxhavelaar.nl/english ▾

You are on the website of the **Max Havelaar** Foundation. **Max Havelaar** is the independent campaigning and certification organization behind the Fairtrade Mark.

[Max Havelaar: Home](#)

www.maxhavelaar.nl/ ▾ [Translate this page](#)

Keurmerk voor eerlijke handel. Maatschappelijk verantwoord ondernemen op basis van een duurzaam inkoopbeleid. Naast informatie over de organisatie ook ...

[Images for "max havelaar"](#) - [Report images](#)

See results about



[Max Havelaar](#)
Book by Multatuli
Max Havelaar: Or the Coffee Auctions of the ...



[Max Havelaar-Stiftung Schweiz](#)
The Max Havelaar Foundation is a ...



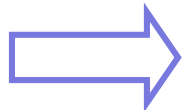
[Association Max Havelaar...](#)
Association Max Havelaar France is the French member of FLO ...



[Stichting Max Havelaar Company](#)
Stichting Max Havelaar is the Dutch member of ...



[Max Havelaar](#)
1976 Film
Max Havelaar is a 1976 Dutch drama film ...



Cached
 Similar
 Share



Web

Images

Maps

Shopping

More ▾

Search tools



36 results (0.53 seconds)

[Max Havelaar: Home](#)www.maxhavelaar.be/ ▾ [Translate this page](#)

Le label du commerce équitable, label octroyé uniquement aux produits qui répondent aux critères internationaux du commerce équitable (conditions de ...

[Accueil - © la Plate-Forme pour le Commerce Équitable](#)www.commerceequitable.org/ ▾ [Translate this page](#)

La Plate-Forme pour le Commerce Équitable (PFCE) est un collectif national de concertation et de représentation d'acteurs de commerce équitable français.

[Artisans du Monde - Accueil](#)www.artisansdumonde.org/ ▾ [Translate this page](#)

Réseau de distribution associatif et militant, Artisans du Monde défend une vision engagée du commerce équitable. Vendre des produits alimentaires et ...

[Ethiquable | Entreprise coopérative, citoyenne et solidaire](#)www.ethiquable.coop/ ▾ [Translate this page](#)

Entreprise coopérative, SCOP, qui propose une gamme de produits du commerce équitable, labellisée par **Max Havelaar** (café, thé, sucre, chocolat, jus de fruit.

[Max Havelaar \(association\) - Wikipédia](#)[fr.wikipedia.org/wiki/Max_Havelaar_\(association\)](http://fr.wikipedia.org/wiki/Max_Havelaar_(association)) ▾ [Translate this page](#)

Max Havelaar [maks 'avlɑːr!] est une branche de l'association internationale Fairtrade Labelling Organizations International (FLO). **Max Havelaar** est ...

[Le Marché Citoyen - annuaire bio, équitable, solidaire, local](#)www.lemarchecitoyen.net/ ▾ [Translate this page](#)

Trouvez les commerces et services écologiques et bio, du commerce équitable ou solidaires, près de chez vous et dans tous les secteurs de consommation.

[AVSF: Association de solidarité internationale pour soutenir l...](#)www.avsf.org/ ▾ [Translate this page](#)

*Franstalige
uitgangspagina
geeft veel Franstalige
resultaten, omdat die
veel onderlinge
linkrelaties hebben*

*oorspronkelijke
zoekterm hoeft er
niet in voor te komen*

terugkoppeling

gebruiker klikt bij relevante hit op *“more like this” / “similar”*

- computer zoekt op grond van daarin aanwezige termen of patronen naar daarop lijkende documenten

bijv.: *Google*
 Autonomy

gebruiker markeert relevante hits

- zoektermen of patronen die in die documenten voorkomen krijgen hoger gewicht
 - bij toekomstige zoekacties krijgen documenten met die termen of patronen een hogere berekende relevantie

bijv.: *Autonomy*
 personal Google

*door te registreren welke
zoekresultaten wel/niet
worden aangeklikt*

terugkoppeling

gebruiker klikt bij relevante hit op *"more like this" / "similar"*

- computer zoekt op grond van daarin aanwezige termen of patronen naar daarop lijkende documenten

bijv.: *Google*

Autonomy

gebruiker merkt relevante hits

- zoektermen of patronen die in die documenten voorkomen krijgen hoger gewicht

→ bij toekomstige zoekacties krijgen documenten met die termen of patronen een hogere berekende relevantie

bijv.: *Autonomy*

personal Google

**kiezen van juiste betekenis of context
(door gelijkenis of door
personalisatie van relevance ranking)
→ betere precisie**

semantisch zoeken

Google

what is the distance from the earth to the moon

Web Shopping News Images Videos More Search tools


wat krijgt u liever?

dit → About 42,200,000 results (0.69 seconds)

of

dat →

238,900 miles
(384,400 km)
Moon, Distance to Earth



Feedback

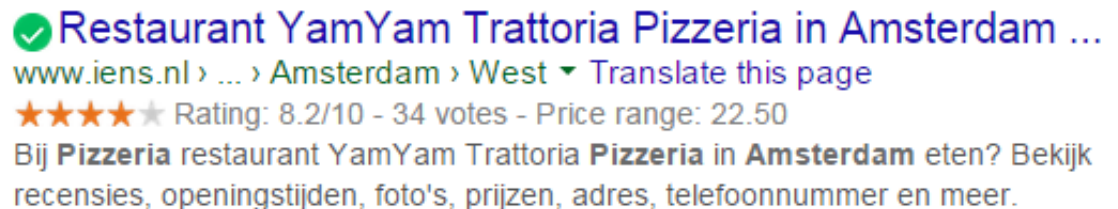
Google geeft steeds vaker concrete antwoorden ("direct answers") op onze zoekvragen

"liever een antwoord dan 10 blauwe links"

semantisch zoeken

wat moet computer kunnen om antwoord te geven?

1. begrijpen wat de zoeker bedoelt en diens intentie is
 - op basis van diens locatie
 - op basis van gebruikte apparaat
 - op basis van eerder zoekgedrag
 - op basis van in vraag herkende entiteiten
 - op basis van natuurlijke taal interpretatie (Google RankBrain)
2. weten wat het antwoord is ("het staat klaar")
 - zoals met Google Knowledge Graph
3. begrijpen wat documenten "bedoelen", zodat daaruit antwoorden kunnen worden afgeleid
 - bijv. uit schema.org codering
4. in *tien blauwe links* relevante context tonen
 - rich snippets /
schema.org



✓ Restaurant YamYam Trattoria Pizzeria in Amsterdam ...
www.iens.nl › ... › Amsterdam › West ▼ Translate this page
★★★★★ Rating: 8.2/10 - 34 votes - Price range: 22.50
Bij Pizzeria restaurant YamYam Trattoria Pizzeria in Amsterdam eten? Bekijk recensies, openingstijden, foto's, prijzen, adres, telefoonnummer en meer.