

uogin

eric sieverts

LexisNexis®

Google

bing

YAHOO!

OvidSP

Onder de motorkap:
techniek van zoeksystemen

techniek van zoeksystemen

- hoe ze werken
 - velden
 - indexen
- basis-functionaliteit voor de gebruiker
 - index bekijken
 - zoeken
 - trunkeren
 - combineren
 - best match
 - filteren
 - resultaten tonen
 - zoekgeschiedenis
- het zoekproces


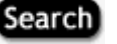



zoeksystemen

bijna alle zoeksystemen zijn gebaseerd op dezelfde basistechnieken en bieden in principe gelijksoortige basisfuncties

- web-zoekmachines
- betaalde zoeksystemen (*ovid, lexisnexis, ebSCO, csa, proquest, stn, ...*)
- web-databases (*pubmed, imdb, eric, ...*)
- discovery tools
- bibliotheekcatalogi

waarop nadruk ligt, hoe je ze "bedient" en hoe ze resultaten presenteren, kan wel sterk verschillen

soorten acties

	ovid zoekstelsel	webzoekmachines
zoeken	perception 	perception 
in index bladeren	percep 	<i>niet</i>
booleaans combineren van termen of van eerdere resultaten	perception and time perception or perceived perception not visual (1 or 2) and 3 not 4	perception time perception OR perceived perception -visual <i>niet</i>
getrunceerd zoeken	percept* organi?ation	<i>niet</i> <i>niet</i>
op specifiek veld zoeken	perception.ti. peereboom*.au.	intitle:perception (bijv. Google, Yahoo) <i>niet</i>
nabijheidszoeken (maximum woordafstand, in zelfde zin, ...)	time adj perception time adj3 perception <i>niet</i>	"time perception" time AROUND(3) perception (Google) time NEAR perception (Yahoo, Exalead)
zoekresultaat tonen in ander format	(<i>automatisch</i>)  Customize Display	(<i>automatisch</i>) <i>niet</i>
zoekgeschiedenis		<i>niet</i> (google history wel een beetje)

waarin zoekt zoekstysteem?

in (bijna) alle digitaal beschikbare tekst

- bij webzoekmachine in alle zichtbare (*en soms in deel van onzichtbare*) tekst van webpagina's
- structuur? <title> / [link tekst](#) / url / ...
- bij bibliografisch bestand in alle onderdelen
- met veldstructuur! >> *voorbeeld*

01064258 INSPEC Abstract Number: A77049149

Title: Summer weather in south-east England, 54 BC

Author(s): Meaden, G.T.

Author Affiliation: Tornado & Storm Res. Organisation, Trowbridge, UK

Journal: Weather vol.32, no.1 p.33-5

Publication Date: Jan. 1977

CODEN: WTHRAL ISSN: 0043-1656

Language: English

Abstract: The meteorological aspects of Julius Caesar's expedition to Britain in 54 BC are discussed. It is shown that the invasion was delayed by a persistent north-west wind, and did not take place until about July 6. The ships reached land near Deal, where they were badly damaged by a storm on the night of 8-9 July. The evacuation, in mid-September, was also hampered by the continuing northwest winds. The occurrence of these northwest winds, and of a drought in Gaul, suggests a prolonged northward extension of the Azores high, with the main anticyclone lying west of the British Isles. (1 Refs)

Descriptors: atmospheric pressure and density; history; meteorology; rain ; storms; wind

Identifiers: 54 BC; storm; drought; Gaul; Azores high; SE.England summer weather; July to September 54 BC; Julius Caesar; NW.winds; N.Atlantic anticyclone

Class Codes: A9260G (Winds and their effects); A9260Q (Storms); A9330G (Europe)

titel
auteurs
werkadres
bron
datum
tijdschr
taal
abstract

trefwoorden
(gecontr.)
trefwoorden
(vrije)

classificatie

indexen

zoeken gebeurt via alfabetische indexen op de inhoud (alle woorden) van in het systeem aanwezige tekst

wat zijn de “termen” die in de index worden gezet?

- woord-index *losse woorden uit de tekst*
- "phrase"-index" *bevat "zinnetjes" uit de tekst
bijv. samengestelde trefwoorden
volledige auteursnamen*

wat wordt met “velden” gedaan?

- afzonderlijke indexen per veld
- één totale index (*met daarin ook veld-info*)

Voorbeeldbestandje met 10 tekstdocumenten

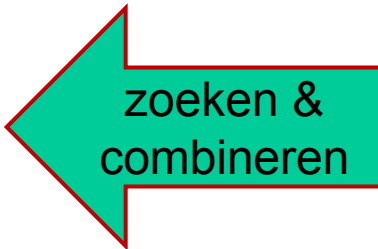
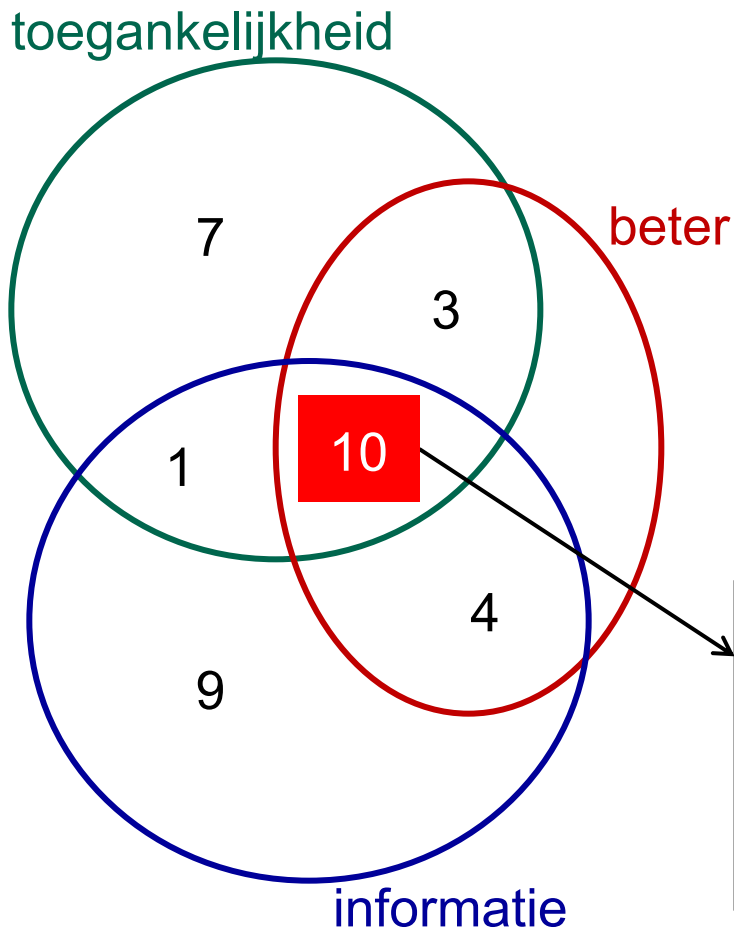
1. Voor de toegankelijkheid van deze informatie gebruiken we een thesaurus
2. Voor browsen gebruiken we een classificatie
3. Zoeksystemen worden beter als we een thesaurus gebruiken voor toegankelijkheid
4. Je kunt informatie beter vinden als in zoeksystemen postcoördinatie wordt toegepast
5. In een classificatie worden aspecten van de inhoud van een document precoördinatief gekoppeld
6. Als een thesaurus wordt toegepast zijn aspecten van de inhoud van een document niet precoördinatief gekoppeld
7. Zoeksystemen die voor toegankelijkheid een thesaurus gebruiken, worden vaak toegepast
8. Postcoördinatie van zoeksystemen maakt dat je niet kunt browsen
9. De inhoud van informatie in een document is gekoppeld aan de postcoördinatie van een thesaurus
10. Hoe beter de toegankelijkheid van informatie, hoe beter we het vinden en hoe beter we de inhoud gebruiken



index

aan	9
als	3,4,6
aspecten	5,6
beter	3,4,10
browsen	2,8
classificatie	2,5
dat	8
de	1,5,6,9,10
deze	1
die	7
document	5,6,9
een	1,2,3,5,6,7,9
en	10
gebruiken	1,2,3,7,10
gekoppeld	5,6,9
het	10
hoe	10
in	4,5,9
informatie	1,4,9,10
inhoud	5,6,9,10
is	9
je	4,8
kunt	4,8
maakt	8
niet	6,8
postcoördinatie	4,8,9
precoördinatief	5,6
thesaurus	1,3,6,7,9
toegankelijkheid	1,3,7,10
toegepast	4,6,7
vaak	7
van	1,5,6,8,9,10
vinden	4,10
voor	1,2,3
we	1,2,3,10
worden	3,5,7
wordt	4,6
zijn	6
zoeksystemen	3,4,7,8

beter AND toegankelijkheid AND informatie



10. Hoe *beter* de *toegankelijkheid* van *informatie*, hoe *beter* we het vinden en hoe *beter* we de *inhoud* gebruiken

aan	9
als	3,4,6
aspecten	5,6
beter	3,4,10
browsen	2,8
classificatie	2,5
dat	8
de	1,5,6,9,10
deze	1
die	7
document	5,6,9
een	1,2,3,5,6,7,9
en	10
gebruiken	1,2,3,7,10
gekoppeld	5,6,9
het	10
hoe	10
in	4,5,9
informatie	1,4,9,10
inhoud	5,6,9,10
is	9
je	4,8
kunt	4,8
maakt	8
niet	6,8
postcoördinatie	4,8,9
precoördinatie	5,6
thesaurus	1,3,6,7,9
toegankelijkheid	1,3,7,10
toegepast	4,6,7
vaak	7
van	1,5,6,8,9,10
vinden	4,10
voor	1,2,3
we	1,2,3,10
worden	3,5,7
wordt	4,6
zijn	6
zoeksystemen	3,4,7,8



index raadplegen

- bij web-zoekmachines is de index niet zichtbaar voor de gebruiker
- bij andere zoeksystemen is die index meestal wel te raadplegen, bijv. om schrijfwijze, woordvorm of frequentie van voorkomen van zoektermen te bekijken

expand weather

Ref	Items	Index-term
E1	2	WEAST
E2	2	WEATE
E3	1542	WEATHER — los woord
E4	1	WEATHER BROADCASTS
E5	25	WEATHER FORECASTING — phrase
E6	2	WEATHER MODIFICATION
E7	4	WEATHER REPORTS
E8	8	WEATHER SATELLITES
E9	2	WEATHER VANES
E10	2	WEATHERALD
E11	5	WEATHERALL
E12	1	WEATHERBEE
E13	1	WEATHERBY
E14	1	WEATHERCASTS
E15	1	WEATHERDISC
E16	12	WEATHERED
E17	27	WEATHERFORD
E18	1	WEATHERFORD HAMMOND MOUNTAIN COLLECTION
E19	1	WEATHERFORD PUBLIC SCHOOL DISTRICT OK
E20	1	WEATHERGLASS
E21	3	WEATHERHEAD
E22	90	WEATHERING
E23	27	WEATHERIZATION
E24	1	WEATHERIZATON

zoeken

- zoeken:
 - meestal alleen maar een woord intikken,
 - soms zoekknop om zoeken te starten,
 - zelden nog zoek-opdracht nodig
- zoeken op een woord geeft als resultaat een verzameling (set) van alle documenten (records) waarin dat woord voorkomt
- systeem meldt meestal hoeveel is gevonden
(bij zoekmachines aantallen zeer onbetrouwbaar)

Set	Term Searched	Items
S1	RETRIEVAL	57349
S2	DATABASE	94522

zoeken

- zoeken op een woord:

weather

- specifiek zoeken: specifiek veld

*weather.ti.
intitle:weather
AU=jansen, a.*

NB: sterk uiteenlopende syntax

- ruimer zoeken: trunkeren

*informat**

trunkeren (en maskeren)

librar*	librarian librarians librarianship libraries library	cat*	cat catacomb catalog catalytic cataract catastrophe catechism category catering caterpillar catharsis cathode catholic
---------	--	------	---

trunkeren (en maskeren)

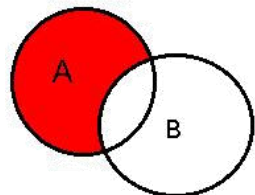
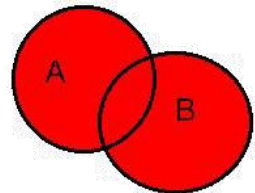
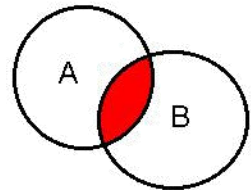
cat+	cat cats catz
cat+++	cat cater cats cattle
wom?	woman women
polari?ation	polarisation polarization

- NB:*
- elk zoekstelsel gebruikt zijn eigen trunkatiesymbolen
 - webzoekmachines kunnen helemaal NIET trunkeren, maar zoeken wel meestal automatisch al op enkel- en meervoud

combineren

gebruik van booleaanse operatoren om woorden (of sets) te koppelen

- AND beide woorden gelijktijdig aanwezig, aan beide criteria voldaan
computer AND bibliotheek
- OR minstens één van de woorden aanwezig, aan minstens één criterium voldaan
fiets OR rijwiel
- NOT woorden/resultaten uitsluiten
greenhouse NOT climatic change



combineren

- combineren van woorden (*ook bij webzoekmachines*)
library AND (web OR internet) AND (use OR using)
- combineren van sets (*niet bij webzoekmachines*)
(#1 OR #2 OR #3) AND (#5 OR #6) NOT #4
- specifiekere AND-combinatie:
 - woorden binnen bepaalde afstand
electronic ADJ4 publishing
electronic NEAR publishing (*zelfs bij Yahoo, Bing, Exalead*)
 - woorden in zelfde zin
library SAME computer

best match

"best match" & "relevance ranking"

- gebruiker tikt reeks zoektermen in (of plakt heel tekstfragment in zoekvenster)
- computer vindt documenten die daar het meest op lijken (het "best" mee overeenkomen/matchen) en presenteert die in volgorde van berekende "relevantie"

maar hoe doet ie dat ?

- *standaard methode bij webzoekmachines*
- *ook mogelijk bij sommige andere zoeksystemen*

OVID

<p>Basic Search (Natural Language)</p>	<p>Natural language search: Works best when entering a phrase or question about the topic of your research. Geared towards novice or occasional researchers but also a powerful tool for professionals that quickly need to find relevant results. e.g.: <i>use of Ritalin for adults with ADHD</i></p>
<p>Ovid Advanced Search</p>	<p>Search on keyword, title author or journal. Will try to map your search term directly to the correct tree or thesaurus term when possible. Supports Boolean search terms and command line syntax searching, e.g. <i>cats.ti,ab. not dogs.ti,ab.</i></p>
<p>Multi Field Search</p>	<p>Easily create Boolean searches. Use the dropdown to select specific fields and choose from the Boolean operators "Not", "Or", "And" to build your precision search.</p>

EBSCO



Searching: **SocINDEX** | [Choose Databases »](#)

therapy alcoholism alcohol abuse adolescents youth

[Search](#) [Clear](#) [?](#)

[Search Options](#) |
 [Basic Search](#) |
 [Advanced Search](#) |
 [Visual Search](#) |
 [Search History/Alerts](#) |
 [Preferences »](#)

Search Options [Reset](#)

Search modes [?](#)

- Boolean/Phrase
- Find all my search terms
- Find any of my search terms
- SmartText Searching** [Hint](#)

Apply related words

SmartText Searching [?](#) [X](#)

Enter as much text for your search as you want - a phrase, a sentence, paragraph, or even whole pages. This mode searches EBSCO Databases only.

[Close](#)

Limit your results

Linked Full Text (Peer

best match

"best match" & "relevance ranking"

wat "matcht" het best?

- hoe meer termen aanwezig hoe beter
*(alleen toegepast bij "kleinere" zoeksystemen;
bij webzoekmachines nu standaard: "all of the terms")*
- document belangrijker als zoekwoorden in titel of heading
- document belangrijker als zoekwoorden herhaald worden
- document belangrijker als zoekwoorden dicht bij elkaar
- zeldzame zoekwoorden belangrijker
- webpagina's belangrijker als er veel naar gelinkt wordt
- . . .

filteren

de meeste zoeksystemen bieden mogelijkheden zoekresultaten te filteren op formele criteria

- publicatiedatum
- documentsoort
- taal van document
- doelgroep
- proefpersonen
- . . .

[details afhankelijk van vakgebied]

meestal gerealiseerd door AND relatie met een gestandaardiseerd kenmerk

Article types

Clinical Trial
 Comparative Study
 Dataset
 Journal Article
 Randomized Controlled Trial
 Review
 Customize ...

Text availability

Abstract
 Free full text
 Full text

Publication dates

5 years
 10 years

Custom range...

Species

Humans
 Other Animals

Sex

Female
 Male

Ages

Child: birth-18 years
 Infant: birth-23 months
 Adult: 19+ years
 Adult: 19-44 years
 Aged: 65+ years
 Customize ...

[Clear all](#)

[Show additional filters](#)

resultaten tonen

opvragen van bepaalde gevonden gegevens,
uit bepaald zoekresultaat, in bepaalde presentatievorm

- bepaald zoekresultaat *set-nummer*
- bepaalde gevonden gegevens *items*
- bepaalde presentatie *format / velden*

*veel zoeksystemen laten na elke zoekactie automatisch
meteen al de eerste 10 gevonden titels zien
(bij zoekmachines kan dat zelfs niet anders)*

... en dan ook zo downloaden of printen

zoekgeschiedenis

overzicht van gebruikte zoekopdrachten en daarmee verkregen resultaten (aantallen)

geheugenopfrisser bij maken van nieuwe combinaties of bij opvragen van resultaten voor printen of downloaden

Set	Term Searched	Items
S1	RETRIEVAL	57349
S2	DATABASE?	121053
S3	DATABANK?	1335
S4	ADVANCES OR TRENDS	88199
S5	S1 AND (S2 OR S3) AND S4	487

*bij sommige web-zoekmachines kunnen gebruikte zoekvragen worden bewaard en weer opgevraagd, maar **geen** te combineren sets [[>> google](#)]*



Goog



Nov 17, 2005

gaat ook ver terug in de tijd
(en je kunt erin zoeken)



Searched for computer

9:10pm



12-computer.jpg



12-computer.jpg

top1.jpg

See 1 more



Searched for computer icons

9:06pm



Searched for computers

9:06pm



Searched for document

8:58pm

document.gif



Searched for trechter

8:51pm



Searched for funnel

8:51pm

alleen

ke

lijkt

zoekproces - praktijk

- beginnen *te gebruiken zoekstelsel kiezen*
(eventueel) bestand kiezen
- zoekactie *in index kijken*
zoeken op termen
combineren
tonen van resultaten
beoordelen van resultaten
bijstellen van zoekactie
zoeken op termen
combineren
tonen van resultaten

.....
inperken op secundaire criteria
afdrukken of downloaden van resultaten

