



K O N I N K L I J K E N E D E R L A N D S E
A K A D E M I E V A N W E T E N S C H A P P E N

Het einde van het traditionele woordenboek

van der Sijs, N.

published in

NRC Next

2010

document version

Publisher's PDF, also known as Version of record

[Link to publication in KNAW Research Portal](#)

citation for published version (APA)

van der Sijs, N. (2010). Het einde van het traditionele woordenboek. *NRC Next*, 18-18.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

pure@knaw.nl

NRC.NEXT

11 maart 2010 donderdag

Betekent 'indiaan' een sterrenbeeld?;

Ja, wel volgens het verwarrende woordenboek van Google

BYLINE: Nicoline van der Sijs

SECTION: OPINIE; Blz. 18

LENGTH: 1028 woorden

SAMENVATTING

Google Woordenboek wordt een alternatief voor het traditionele woordenboek.

Maar houd je Van Dale nog maar even in de kast. Google levert een hoop onzin op.

VOLLEDIGE TEKST:

De nieuwe dienst Google Woordenboek biedt gratis moderne 'eentalige' woordenboeken aan voor allerlei talen, waaronder het Nederlands, en vertaalwoordenboeken naar of vanuit het Engels.

Vernieuwend is de wijze waarop de informatie verzameld wordt. De andere woordenboeken en vertaalwoordenboeken die op internet zijn te vinden, zijn gewoon de oude 'papieren' woordenboeken, maar dan in digitale vorm op internet. De woordenboeken van Google zijn het resultaat van programmatuur die het hele internet afzoekt op woorden en betekenissen.

De ontwikkeling van Google Woordenboek sluit naadloos aan bij de open access-filosofie, die ontegenzeggelijk de toekomst heeft. Die filosofie houdt in dat culturele en wetenschappelijke informatie algemeen beschikbaar moet worden gesteld op internet.

De achterliggende zoektechniek is voor de gebruiker niet zichtbaar, maar uit de resultaten kan wel het een en ander worden opgemaakt. Google Woordenboek zoekt voor ieder trefwoord websites af waarop definities van woorden worden gegeven en plaatst de gevonden resultaten in een bepaalde hiërarchische volgorde, onder andere naar aantallen 'hits' - precies zoals de zoekfunctie van Google werkt. Daarbij wordt in ieder geval de informatie getoond uit Wikipedia, Wiktionary en eventuele andere woordenboeken of encyclopedieën die

beschikbaar zijn op internet, en voorts uit informatieve websites, zoals die met overheidsinformatie.

Bij het eentalige Engelse woordenboek, waarvan de ontwikkeling momenteel het verst is, wordt bij ieder trefwoord behalve een groot aantal betekenisomschrijvingen ook informatie gegeven over uitspraak (met geluid), woordsoort, verbuiging of vervoeging, woorden met dezelfde of tegengestelde betekenis, voorbeeldzinnen, vaste verbindingen, samenstellingen. Daarnaast zijn soms plaatjes, geluid en filmpjes toegevoegd.

Het feit dat het grootste corpus ter wereld, het internet, gebruikt wordt voor de verzameling en beschrijving van trefwoorden, levert de meest up-to-date informatie op en - mede daardoor - een groot aantal onverwachte gegevens die woordenboekschrijvers gemakkelijk missen. Dat blijkt als we de gegevens in Google Woordenboek vergelijken met die uit vergelijkbare woordenboeken: de papieren Dikke Van Dale en het digitale Algemeen Nederlands Woordenboek (ANW). Het ANW is een digitaal woordenboek van het eigentijdse Nederlands dat wordt vervaardigd op het Instituut voor Nederlandse Lexicologie; het is gebaseerd op een statisch en afgesloten corpus en bevat momenteel ruim 900 woordenboekartikelen. In 2019 moeten dat er 70.000 zijn.

In alle drie deze woordenboeken staat dat Apache een indianenvolk of lid daarvan is. Van Dale vermeldt nog dat het met kleine letter betekent: 'lid van de onderwereld'. Dat vermeldt het ANW niet, maar weer wel dat het de naam van een gevechtshelikopter is. Aan deze informatie voegt Google Woordenboek toe: Apache is een taalnaam, de naam van een Amerikaanse motorfiets, een webserver, een bordspel, een gemeente in de Amerikaanse staat Oklahoma, en de naam van het paard van de beroemde dressuuramazone Isabell Werth.

Google Woordenboek wint het dus nu al, terwijl de Nederlandse versie nog maar net in ontwikkeling is, van representatieve papieren en digitale woordenboeken.

De ontwikkeling van Google Woordenboek luidt het einde in van het papieren eentalige en tweetalige moderne woordenboek, zoveel lijkt wel zeker: zodra deze gratis dienst onder gebruikers algemeen bekend raakt, zullen ze massaal hun dure papieren woordenboeken de deur uit doen en hun abonnementen op digitale woordenboeken opzeggen, zoals ze eerder deden met hun encyclopedieën.

Maar doen ze daar goed aan? Dat de resultaten wel degelijk redactionele beoordeling nodig hebben, blijkt als men bijvoorbeeld 'indiaan' opzoekt. De eerste betekenis is zeer ongebruikelijk: 'onopvallend sterrenbeeld nabij de zuidelijke hemelpool' en de laatste betekenis, afkomstig uit een volksalmanak, is onduidelijk. Computers kunnen wel materiaal voor woordenboeken verschaffen, maar ze maken beslissingen van woordenboekmakers niet overbodig.

Ook om een andere reden lijkt het verstandig om kritisch te blijven tegenover de resultaten van Google Woordenboek, hoe veelbelovend die er ook uitzien. De achterliggende algoritmes van Google zijn niet bekend en de zoekmachine komt de laatste tijd steeds meer onder vuur te liggen omdat ze, vrijwillig of onder dreiging van juridische stappen, informatie verwijderd of blokkeert.

Een zoekmachine die bereid is tot censuur kan beter niet het monopolie op lexicografische producten krijgen. Dat zou de weg openen naar George Orwells politiek correcte Newspeak,

de enige taal ter wereld waarvan de woordenschat en de woordbetekenissen ieder jaar afnemen in plaats van toenemen.

'Indiaan' volgens Google

Als je het woord 'indiaan' opzoekt in Google Woordenboek, krijg je het resultaat hieronder.

Opvallend is dat als eerste een zeer ongebruikelijke en weinig frequente betekenis wordt gegeven. De relevantie van de laatste 'betekenis' is onduidelijk; de link verwijst naar een volksalmanak met 1.881 volksverhalen uit verschillende landen en culturen.

„Indiaan (Indus, afkorting Ind) is een onopvallend sterrenbeeld nabij de zuidelijke hemelpool, tussen rechte klimming 20u25m en 23u25m en tussen declinatie -45° en -75°. Vanaf de breedte van de Benelux is het niet te zien." Bron: [nl.wikipedia.org/wiki/Indiaan_\(sterrenbeeld\)](http://nl.wikipedia.org/wiki/Indiaan_(sterrenbeeld))

„Indianen, de oorspronkelijke bewoners van Amerika *Indiaan (sterrenbeeld), een sterrenbeeld* In het 19e eeuwse Nederland een gebruikelijke aanduiding voor de inwoners van Nederlands-Indië, later kwam 'inlander' in zwang. *El Indjan, een rotsformatie op Curaçao." Bron: nl.wikipedia.org/wiki/Indiaan

„Inwoner van Amerika voor de komst van de Europeanen." Bron: nl.wiktionary.org/wiki/indiaan

„Indianen - De geesten van de nacht, Mani, De aap Lawaaischopper en de maïskolven. Hoe de jaguar overwonnen werd, De schildpad Jabuti en het vuur." Bron: www.beleven.org/verhalen/info/de_slangen_ruilen_de_nacht