Twee concepten van informatie: Absoluut en relatief

prof. Pieter W. Adriaans, Human Computer Studies, UvA, RoboSail
(pietera@science.uva.nl)

Het woord 'informatie' is een verzamelterm, net als de woorden 'energie', 'waarde', 'arbeid' en 'kracht'. Dit soort woorden worden in het dagelijks leven te pas en te onpas gebruikt, soms om vage algemene begrippen aan te duiden, soms in een strikt wiskundig gedefinieerde wetenschappelijke betekenis. Die betekenissen zijn vaak onhelder afgebakend en worden door elkaar gebruikt. Het natuurkundige begrip 'arbeid' is iets totaal anders dan waar Marx in Das Kapital over schreef. Zo zeggen we: "Jan had geen energie meer om zijn huiswerk af te maken", en "De centrale leverde energie voor de hele stad", maar niemand concludeert dat Jan zijn huiswerk niet maakt omdat de stroom is uitgevallen.

De E in de bekende formule van Einstein: $E = mc^2$ staat voor energie, maar ook dat heeft weinig met de 'energie' van Jan te maken. Vaak hebben deze verzameltermen verschillende wiskundige definities die wel het voordeel hebben dat ze exact zijn, maar die slechts een deel van alle mogelijke betekenissen dekken. Hetzelfde geldt voor het concept 'informatie'. In dit artikel zal ik twee betekenissen van het woord 'informatie' uitwerken - relatieve en absolute informatie - en uitleggen wat de formele betekenis ervan is. Deze totaal verschillende noties van informatie worden nogal eens door elkaar gehaald. Als we de definities nauwkeurig bekijken, blijkt dat de studie van het begrip informatie ons met een aantal diepe filosofische vragen confronteert.

Laten we eerst een paar voorbeelden bekijken:
1) Alice heeft van Bob een lot in een loterij gekocht en ze is benieuwd of ze gewonnen heeft. Zij zou graag informatie over het winnende lotnummer hebben.
2) Een encyclopedie bevat veel informatie.
3) Alice is een spion en ze heeft een gecodeerde brief van Bob onderschept. Zij wil graag weten welke informatie die brief bevat.
4) Alice ontvangt een onbekend radiosignaal uit de ruimte. Ze is benieuwd of het signaal ook informatie bevat.
5) Wetenschappers kunnen uit foto's van het gesteente van Mars afleiden dat er vroeger water op Mars geweest moet zijn. Welke informatie speelt daarbij een rol?

We noemen informatie discreet als ze in eindige pakketjes verpakt kan worden. Alle informatie die voor ons relevant is, kan discreet gemaakt worden. Zelfs informatie die voor onze zintuigen continu lijkt, zoals muziek of film, kan tegenwoordig makkelijk digitaal worden opgeslagen. We maken de elektronica gewoon zo snel dat onze arme, langzame, biologische zintuigen het verschil met werkelijk continue informatiestromen in de werkelijkheid niet bemerken. In de wiskunde komen wel niet-discrete vormen van informatie voor, bijvoorbeeld continue verzamelingen reële getallen, maar die zijn voorlopig voor praktische toepassingen nauwelijks van belang. De kwantummechanica leert ons zelfs dat de hele natuur in essentie discreet is. Energiestromen zijn niet continu. Ook de tijd verloopt in kleine stapjes.

De meest fundamentele eenheid van informatie is de bit. Een bit kan men denken als een schakelaar met twee standen: aan of uit, 1 of 0. De hoeveelheid verschillende boodschappen die we met n bits kunnen coderen is 2^n. Rekenen met bits staat gelijk aan rekenen met machten van 2. Een bit is een abstracte notie die op heel veel verschillende manieren in de natuur gerealiseerd kan worden: elektronische schakelingen, kiezelsteentjes in het zand, kralen aan een ketting, kwantumtoestanden, munten of deuren die open of dicht kunnen zijn.

Niet alle systemen in de natuur zijn geschikt om informatie in op te slaan. Alleen stabiele systemen die reversibel zijn, komen daarvoor in aanmerking. Een systeem is reversibel als het twee toestanden kent, zeg A en B, en als het van toestand A over kan gaan in toestand B, en van toestand B weer in A. We willen ook dat die toestanden A en B stabiel zijn. Als er geen energie aan het systeem wordt toegevoegd, moet het blijven zoals het is. Een elektronische schakelaar is een dergelijk systeem, een deur ook. Een voorbeeld van een irreversibele gebeurtenis is het oplossen van melk in een kop koffie. Wat we ook doen, als de melk zich eenmaal regelmatig door de koffie verspreid heeft, blijft dat zo. Nooit kunnen we het systeem weer splitsen in pure koffie en melk, hoewel dat statistisch in principe van zelf zou kunnen gebeuren. Pure energie-neutrale reversibele systemen komen in de natuur niet voor. Volgens de tweede hoofdwet van de thermodynamica, zal er altijd energie nodig zijn om van toestand A via B weer naar A terug te komen. Reversibele systemen zijn in de natuur meer uitzondering dan regel. De meeste processen in de natuur zijn irreversibel: golven op zee, de wind boven land, het oplossen van rook in de lucht, de expansie van het universum zelf.

Het is mooi dat we de bit als eenheid van informatie hebben, maar zonder een interpretatie van die eenheid betekent het niet veel. We moeten manieren hebben om de eenheid toe te passen.

Eén opvatting, die we de relatieve interpretatie van informatie zullen noemen, relateert de hoeveelheid informatie die in een boodschap is vervat aan de waarschijnlijkheid dat die boodschap optreedt. Een boodschap die erg onwaarschijnlijk is, bevat veel informatie. Een boodschap waarvan we bijna zeker weten dat die komt, bevat nauwelijks informatie. Stel dat we de uitdrukking $I_r(p)$ introduceren voor de hoeveelheid relatieve informatie die in boodschap p is vervat en de uitdrukking $P(p)$ voor de waarschijnlijkheid dat p optreedt. Nu gelden de volgende basisregels:

1) Een boodschap die optreedt met waarschijnlijkheid 1 bevat 0 bits aan informatie: als $P(p) = 1$ dan $I_r(p) = 0$.

2) De informatie in twee boodschappen p en q die onafhankelijk van elkaar zijn, is de som van de informatie in de afzonderlijke boodschappen: $I_r(p \ \& \ q) = I_r(p)+I_r(q)$.

Een formule die aan deze basisregels voldoet is $I_r(p) = -\log_2 P(p)$. Check maar:
1) Als $P(p) = 1$ dan geldt $I_r(p) = -\log_2 P(p) = -\log_2 1 = 0$.
2) Als p en q onafhankelijk zijn dan geldt $I_r(p \ \& \ q) = -\log_2 P(p \ \& \ q) = -\log_2 (P(p) \times P(q)) = -\log_2 P(p) - \log_2 P(q) = I_r(p) + I_r(q)$.

Er is een interpretatie van het begrip relatieve informatie die past als een handschoen: die waarin we de relatieve frequentie van een boodschap opvatten als zijn waarschijnlijkheid. Een voorbeeld: toen kranten nog in lood gezet werden, wist iedere drukker dat hij meer versies van de letters e en n nodig had dan van de q en de z. De letters e en n komen veel meer voor dan de letters q en z. Bij Scrabble hebben de q en de z ook een hogere letterwaarde. Als we een kruiswoordraadsel willen oplossen dan hebben we veel meer aan een q of een z dan aan een e. Men zou kunnen zeggen dat een q en een z meer informatie bevatten. Omgekeerd geldt: wie een code-systeem wil ontwerpen (bijvoorbeeld Morse of Braille) doet er goed aan de kortste codes te kiezen voor de meest frequente letters en de langere codes te reserveren voor letters de minder vaak voorkomen. In Braille wordt de letter e door één punt weergegeven. Hier is dus de volkomen natuurlijke interpretatie: Voor boodschappen die vaak voorkomen (en dus weinig informatie bevatten) gebruiken we weinig bits, voor boodschappen die zeldzaam zijn (en dus veel informatie geven) gebruiken we langere codes met meer bits. Deze definitie van het begrip informatie is geïntroduceerd door Claude Shannon. Met Shannon's wiskundige fundering van het begrip informatie kunnen we optimale codesystemen ontwerpen. Stel dat we met enige regelmaat een aantal boodschappen over een communicatie kanaal willen sturen. Met n bits kunnen we 2^n boodschappen coderen. Stel $k=2^n$. Als we k boodschappen hebben, waarvan de waarschijnlijkheid uniform verdeeld is, dan geldt dat we voor elke boodschap n bits moeten gebruiken. Als we meer bits gebruiken, dan zit de informatie te ruim in zijn jasje. Er is dan redundantie. Als de waarschijnlijkheidsverdeling niet uniform is, geldt dat $I_r(p) = -\log_2 P(p)$ de optimale lengte van de code voor boodschap p geeft.

Shannon's definitie is wiskundig simpel en dekt zeker een deel van het concept dat we in het dagelijks leven informatie noemen. Toch zijn er problemen. Stel dat we dezelfde boodschap aan twee personen zenden. Als ze niet dezelfde waarschijnlijkheidsdistributie hanteren, dan bevat de boodschap voor de een meer informatie dan voor de ander. De informatie die vervat is in een boodschap is relatief ten opzichte van de verwachting van de ontvanger. Dat maakt de hoeveelheid informatie die in een boodschap zit subjectief. Op zich zit daar wat in. Een boodschap die voor de een als een verrassing komt zal voor de ander nauwelijks nieuws bevatten. Toch is dat niet in alle gevallen bevredigend. Het betekent dat we eigenlijk niet over informatie zonder meer kunnen spreken. Als iemand zegt: "Deze encyclopedie bevat veel informatie", dan is dat niet bedoeld als een relatief statement. Natuurlijk kan de ene lezer meer leren uit een encyclopedie dan de ander, maar de informatie die in de encyclopedie vervat is lijkt een intrinsieke objectieve kwaliteit. Ook al zouden we de encyclopedie in een kluis sluiten en nooit aan iemand laten zien, het blijft een feit dat hij een bepaalde hoeveelheid informatie bevat. Die hoeveelheid informatie lijkt weinig te maken te hebben met de waarschijnlijkheid dat iemand de encyclopedie als boodschap naar iemand anders zendt. Om kort te gaan: Shannon's definitie is waardevol, maar dekt zeker niet de hele lading van het begrip informatie. Er zou ook een soort absolute notie van informatie moeten zijn, onafhankelijk van de beschouwer. Het blijkt dat een dergelijk concept gedefinieerd kan worden, maar het heeft iets meer voeten in de aarde dan de notie van relatieve informatie.

Laten we eerst eens onderzoeken aan welke voorwaarden zo'n concept van absolute informatie zou moeten voldoen. Stel dat we de encyclopedie opslaan op een CD. Dat kost een bepaalde hoeveelheid, zeg k, bits. Toch is het niet verstandig om te stellen dat onze encyclopedie k bits aan informatie bevat. De volgende korte redenering maakt dat duidelijk. Stel dat we de encyclopedie van het Engels naar het Duits vertalen. Over het algemeen zijn teksten in het Duits wat langer dan Engelse teksten. De Duitse versie van de Encyclopedie zal dus meer bits omvatten, maar het is natuurlijk niet zo dat ze ook meer informatie bevat. We zouden op zijn minst een notie van informatie willen hebben die taalonafhankelijk is. Dan is er nog een reden waarom het niet slim is om de lengte van een boodschap als een maat voor de hoeveelheid informatie in een boodschap te beschouwen. Stel dat ik iemand een boodschap stuur die uit allemaal nullen bestaat: "00000…". Ongeacht de lengte van de boodschap kunnen we niet zeggen dat een dergelijke boodschap veel informatie bevat. Verder zouden we graag willen dat de informatie in twee onafhankelijke boodschappen p en q de som is van informatie in p en in q. Om kort te gaan, stel dat we de uitdrukking $I_a(p)$ gebruiken voor de absolute informatie in p dan zouden we de volgende eisen willen stellen:

"Een beter begrip van de notie van informatie is een van de fundamentele uitdagingen van de hedendaagse wetenschap."

1) $I_a(p)$ zou onafhankelijk moeten zijn van de taal waarin p geformuleerd is.

2) $I_a(p)$ zou de interne complexiteit van p moeten representeren, d.w.z. lange simpele boodschappen zouden minder informatie moeten bevatten dan korte complexe boodschappen.
3) Als p en q onafhankelijk zijn dan zou moeten gelden:
$I_a(p \ \& \ q) = I_a(p) + I_a(q)$

Er is een formele definitie van informatie die in zekere mate aan deze eisen voldoet, in de vorm van de zogenaamde prefix Kolmogorov-complexiteit, maar deze definitie is niet zo elegant als die van Shannon.

Het basisidee is gebaseerd op een concept van Turing. Een Turing-machine is een abstracte machine die bewerkingen op binaire strings kan uitvoeren. Turing bewees dat er zogenaamde universele Turing-machines bestaan. Dat zijn computers die in staat zijn alle andere computers te emuleren. Stel dat we een programma p op een computer C willen laten draaien. We nemen een universele Turing-machine TU en we voeren eerst een beschrijving van C in, zodat TU zich als C gaat gedragen Daarna voeren we programma p uit. Naast dit positieve resultaat (het bestaan van universele Turing-machines) bewees Turing nog een belangrijk negatief resultaat: er bestaat geen programma dat voor alle combinaties van programma's p en computers C kan beslissen of C op invoer p stopt of niet. Soms zullen computers op bepaalde invoer in een oneindige loop raken, maar we kunnen niet in alle gevallen van tevoren beslissen of dit gebeurt. We noemen dit probleem het 'Haltingprobleem'. De Russische wiskundige A.N. Kolmogorov stelde voor om de informatie-inhoud van een binaire string b te definiëren als: de lengte van het kortste programma dat b produceert op een van tevoren geselecteerde universele Turing machine TU. We noemen deze informatiemaat te zijner ere 'Kolmogorov-complexiteit' C(p) van een binaire string p. We spreken ook wel van de algoritmische complexiteit van een string. De definitie van Kolmogorov is niet vrij van problemen. Ik noem er een paar:

1) Kolmogorov-complexiteit is niet constructief. We kunnen haar niet berekenen. Dit is een rechtstreeks gevolg van het Haltingprobleem. Idealiter zouden we voor het vaststellen van de Kolmogorov-complexiteit van een string p alle kortere binaire strings in moeten voeren in TU en moeten checken of ze p als output opleveren. Dit kan echter niet, want sommige programma's zullen nooit stoppen en we kunnen nooit voorspellen welke.
2) Kolmogorov-complexiteit is relatief ten opzichte van de universele Turing-machine die we gekozen hebben. Als we een andere machine TU kiezen, dan verandert de definitie van Kolmogorov-complexiteit. Dit probleem kan ten dele worden ondervangen. Het blijkt dat de verschillende definities van Kolmogorov-complexiteit die we krijgen hoogstens een contante c van elkaar verschillen. Dit ligt voor de hand, alle machines kunnen elkaar immers emuleren. Men kan al een universele Turing-machine definiëren die minder dan 100 bits groot is. Niettemin maakt dit inzicht het nagenoeg onmogelijk om zinvol te spreken over de Kolmogorov-complexiteit van relatief korte strings. De Kolmogorov-versie van absolute informatie is asymptotisch van karakter; Naarmate de informatie complexer is, wordt ze nauwkeuriger.
3) Kolmogorov-complexiteit is niet additief. Stel dat we twee strings hebben, p en q, en twee programma's p' en q' die de strings genereren. Er is geen garantie dat de concatenatie $p' \bullet q'$ van p' en q' ook een valide programma is dat $p \bullet q$ genereert.

Om dit laatste probleem te ondervangen, gebruiken we de zogenaamde prefix-vrije variant van Kolmogorov-complexiteit, aangegeven door $K(p)$. Een verzameling strings is prefix-vrij als geen enkele string een prefix is van een andere. Het is goed mogelijk om de verzameling van eindige binaire strings prefix-vrij te hercoderen. We krijgen dan een oneindige verzameling van programma's die geen van alle prefixen van een ander zijn. Het voordeel is dat we twee programma's nu eenvoudig aan elkaar kunnen plakken. De machine kan zelf beslissen wanneer de input gesplitst moet worden. Een nadeel is dat hierdoor de programma's wat langer worden. Het is heel goed mogelijk dat $K(p)$ uiteindelijk groter is dan de lengte van p. Niettemin voldoet de definitie $I_a(p) = K(p)$ in zekere mate aan de eisen die we hebben gesteld:

1) $I_a(p) = K(p)$ is met inachtneming van een constante c onafhankelijk van de taal waarin p geformuleerd is.
2) $I_a(p) = K(p)$ representeert de interne complexiteit van p. Een lange simpele string heeft een korter programma dan een korte complexe string.
3) Als p en q onafhankelijk zijn dan geldt:
$I_a(p \bullet q) = K(p \bullet q) < K(p)+K(q) = I_a(p) + I_a(q)$

De reden voor het ongelijkheidsteken in $K(p \bullet q) < K(p)+K(q)$ is dat het altijd mogelijk is dat er voor $p \bullet q$ toevallig een korter programma is dan $K(p)+K(q)$. Een simpel voorbeeld maakt dat duidelijk. Stel, we hebben een string $k = p \bullet q$. Om p en q uit k te halen moeten we informatie toevoegen: een getal dat aangeeft waar k gesplitst moet worden. Als k zelf een lage complexiteit heeft en vrij lang is dan kan het heel goed zijn dat
$K(p \bullet q) < K(p)+K(q)$.

De definitie $I_a(p) = K(p)$ helpt ons een eind in de goede richting, maar is verre van ideaal. $K(p)$ is niet constructief, we kunnen het niet berekenen. $K(p)$ is asymptotisch, het zegt weinig zinvols over korte strings. $K(p)$ is niet volledig additief, als we strings concateneren kunnen we informatie verliezen. Niettemin kunnen we prefix-Kolmogorov wel gebruiken. We kunnen het niet exact meten, maar er wel mee rekenen. De natuur zit vol met systemen die we slechts kunnen benaderen. Dus dát hoeft ons niet te weerhouden. We kunnen nu ook de vraag beantwoorden over de hoeveelheid informatie in de encyclopedie: het is de prefix Kolmogorov-complexiteit van de encyclopedie. Een encyclopedie die meer informatie bevat zal een hogere Kolmogorov-complexiteit hebben.

Een interessante consequentie van de definitie van Kolmogorov-complexiteit is dat random strings de meeste informatie bevatten. Om het anders te zeggen: een signaal dat uit niets anders dan ruis bestaat bevat meer informatie dan een signaal met een duidelijke

"Het lijkt vreemd, maar een signaal dat uit niets anders dan ruis bestaat bevat meer informatie dan een signaal met een duidelijke structuur."

structuur. Dit lijkt vreemd. We hebben niet het gevoel dat een TV kanaal met alleen maar sneeuw meer informatie geeft dan bijvoorbeeld CNN. Toch is dit zo. Een eenvoudig experiment illustreert dit. Neem een bak met lego-steentjes. We vragen aan proefpersoon om een configuratie van lego-steentjes te onthouden en daarna exact te reproduceren. We kunnen twee dingen met de lego-steentjes doen: ze willekeurig uitstorten over de grond, of er een huisje van bouwen. In het eerste geval is het veel moeilijker voor de proefpersoon om de taak te volbrengen dan in het tweede geval. Het kost meer moeite alle willekeurige posities van de steentjes te onthouden dan om het ontwerp van een huisje te reproduceren. Om het anders te zeggen: chaos bevat meer informatie dan orde. Toch zijn mensen biologisch niet erg goed in het exact reproduceren van chaotische structuren. Wij vinden orde over het algemeen interessanter dan chaos. De reden lijkt te zijn dat chaos een natuurlijk eindstadium van alle processen in de natuur is. De tweede wet van thermodynamica zegt dat alle natuurlijke systemen uiteindelijk een stabiele toestand van maximale wanorde zullen bereiken. In een dergelijke toestand bevat het systeem geen vrije energie meer. Je kunt er niets meer mee doen. Zo'n toestand van maximale wanorde (die gelijkstaat aan een maximale hoeveelheid informatie) is niet interessant voor ons. Het is afval waar we geen energie/informatie meer aan kunnen onttrekken. Mensen lijken biologisch zo geprogrammeerd te zijn dat ze vooral interesse hebben in systemen met een relatief lage Komogorov-complexiteit, ofwel comprimeerbare systemen. Dit ligt voor de hand. Als chaos een natuurlijke staat van een systeem is dan is orde iets bijzonders. Iets moet de orde veroorzaakt hebben. Neem het voorbeeld van de lego-steentjes. Het is extreem onwaarschijnlijk dat de lego-steentjes zich vanzelf tot een huisje organiseren. Het maken van een huisje kost meer energie dan het willekeurig over de vloer uitstorten van de steentjes. In een universum waar alles als vanzelf complexer wordt, kost het reduceren van complexiteit energie (of in een formulering die mij als docent meer aanstaat: het handhaven van orde kost energie). In een dergelijke wereld zijn systemen met een relatief lage complexiteit interessanter dan systemen met maximale complexiteit.

Nu we de twee concepten van informatie nader uitgewerkt hebben, is het instructief te kijken naar de voorbeelden uit het begin. Het voorbeeld van het lot uit loterij kan in termen van relatieve informatie worden geanalyseerd. Als er k loten verkocht zijn en er is één prijs, dan is de kans op een prijs $1/k$ en bevat de boodschap over het winnende lotnummer $\log_2 k$ bits. Het geval van de encyclopedie is al in tekst besproken als iets dat meer in termen van absolute informatie geïnterpreteerd dient te worden. Het ontcijferen van gecodeerde brieven is typisch iets waar de notie van relatieve informatie goede diensten kan bewijzen. Het geval van het onbekende radiosignaal uit de ruimte ligt anders. Bij een signaal waarvan we weten dat het een gecodeerde boodschap bevat, hebben we een zekere verwachting over de distributie van de tekens. Bij een signaal uit de ruimte niet. We kunnen het daarom alleen interpreteren in termen van absolute informatie. De meest voor de hand liggende interpretatie is dat het signaal uit ruis bestaat. Als het echter structuur vertoont, dan is dat een indicatie dat er meer aan de hand is. Een zelfde situatie vinden we bij de analyse van de foto's van Mars. De a priori verwachting is een ongeordende chaotische bodemstructuur. Als er structuren gevonden worden die typisch zijn voor de activiteit van water, dan is die conclusie meer voor de hand liggend dan de veronderstelling dat die structuren bij toeval ontstaan zijn.

We hebben nu twee formele betekenissen van het begrip informatie belicht: het concept van relatieve informatie dat gebaseerd is op de waarschijnlijkheid van de boodschap en het concept van absolute informatie dat geassocieerd is met het kortste programma dat een boodschap genereert op een universele Turing-machine. De definitie van relatieve informatie is elegant, de definitie van absolute informatie is rafelig. Op het eerste gezicht lijkt het teleurstellend dat de notie 'absolute informatie' niet netter gedefinieerd kan worden. Wie er even over nadenkt, realiseert zich direct dat we niet veel meer mogen verwachten. Een efficiënte constructieve definitie van de notie van absolute informatie zou alle heuristiek overbodig maken. Het zou de heilige graal van de empirische wetenschap zijn. Om een verschijnsel wetenschappelijk te verklaren zouden we niets méér hoeven te doen dan alle meetgegevens betreffende een bepaald verschijnsel in de computer in te voeren en aan de computer te vragen om de absolute informatie uit de data-set te halen. Om een geheimschrift te ontcijferen zouden we slechts naar de absolute informatie in de boodschap hoeven te zoeken. De cryptografie zou overbodig geworden zijn. Absolute informatie is naar de aard van haar natuur iets dat alleen maar benaderd kan worden. Er is geen vaste universele methode om absolute informatie uit een boodschap te filteren. De hele geschiedenis van de wetenschap kan gezien worden als een strijd voor het vinden van methoden om absolute informatie uit verschillende soorten datasets te halen. Informatie mag dan een concept zijn dat we in het dagelijks leven nagenoeg gedachteloos gebruiken, als we dieper doorvragen blijkt dat de vraag naar de essentie van informatie verre van triviaal is. Een beter begrip van de notie van informatie is een van de fundamentele uitdagingen van de hedendaagse wetenschap. De vraag naar de essentie van informatie raakt aan de meest uiteenlopende wetenschapsgebieden: de informatica, de wiskunde, de natuurkunde en de biologie. Er is nog een hoop werk voor de informatici. ø