

## Het vakgebied informatica

Introductie	15
1	Google 16
1.1	Webpagina's 17
1.2	Webcrawlers 19
1.3	Het maken van de index 21
1.4	Het afhandelen van de zoekvraag 23
1.5	Efficiënt omspringen met ruimte en tijd 24
1.6	Google en de wereld 25
2	Het vak informatica 26
2.1	Wat is informatica? 26
2.2	Deelgebieden van informatica 31
2.3	Ethiek en informatica 38
3	Het vak Inleiding Informatica 39
Samenvatting	40
Zelftoets	41
Terugkoppeling	42
- Uitwerking van de opgaven	42
- Uitwerking van de zelftoets	42



Leereenheid 1

## Het vakgebied informatica

### INTRODUCTIE

Deze cursus geeft een inleiding in het vakgebied informatica; het is daarom gepast om te beginnen met de vraag wat informatica eigenlijk is. Helaas is er niet één algemeen geldende definitie; dit komt omdat er nogal uiteenlopende opvattingen bestaan over wat wel en wat niet tot het vakgebied behoort. In paragraaf 2 zullen we de verschillende visies nader bekijken en een indeling in deelgebieden presenteren van het vakgebied, zodat u een beeld krijgt van het soort zaken dat onderzocht en ontwikkeld wordt. Maar eerst willen we een kader scheppen voor die bespreking, aan de hand van een concreet voorbeeld dat we kunnen gebruiken om de deelgebieden mee te illustreren.

Dat doen we met als uitgangspunt een inmiddels alledaagse handeling, namelijk het zoeken van informatie op het internet met behulp van een zoekmachine zoals Google. Wat komt daar zoal bij kijken en hoe slaagt Google erin om gezochte informatie zo onwaarschijnlijk snel tevoorschijn te toveren?

Het zal duidelijk zijn dat het vakgebied informatica essentieel is geweest voor het ontstaan van het internet en zoekmachines zoals Google. Het blijkt dan ook mogelijk om veel deelgebieden van het vakgebied te illustreren met behulp van dit voorbeeld.

#### LEERDOELEN

Na het bestuderen van deze leereenheid wordt verwacht dat u:

- kunt aangeven waarom er geen eensluitende definitie van informatica gegeven kan worden
- kunt aangeven wat bedoeld wordt met de kerninformaticus, de technische informaticus, de bestuurlijke informaticus en de managende informaticus
- vier hoofddomeinen van de informatica kunt noemen en kort kunt aangeven waar die op betrekking hebben
- van elk hoofddomein enkele deelgebieden kunt noemen en kort kunt aangeven waar die op betrekking hebben.

#### *Studeeraanwijzingen*

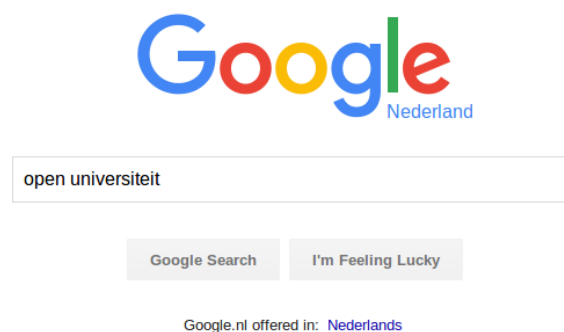
Deze leereenheid is bedoeld om een eerste indruk te krijgen van het vakgebied. Het is vooral een leesleereenheid; het aantal opgaven is tot een minimum beperkt. De studielast van deze leereenheid bedraagt circa 3 uur.

## L E E R K E R N

## 1 Google

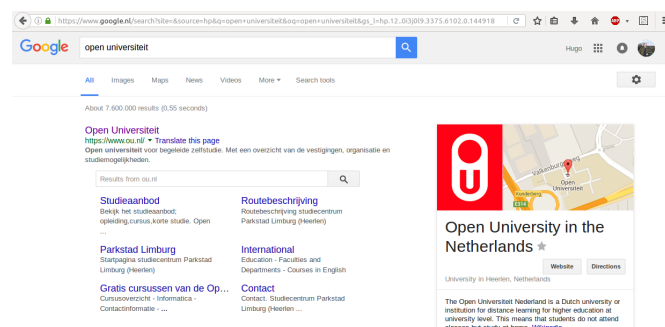
We doen het zo vaak dat het een werkwoord is geworden: googelen, iets opzoeken op 'het internet' met behulp van Google, op dit moment een van de populairste zoekmachines.

Laten we het meteen maar eens proberen: we surfen met behulp van een webbrowser zoals Internet Explorer naar [www.google.nl](http://www.google.nl) en typen 'open universiteit' in (zie figuur 1.1) om informatie over de Open Universiteit te zoeken.



FIGUUR 1.1 Informatie zoeken met Google

Een druk op de knop Google zoeken levert maar liefst zo'n 7 600 000 resultaten (op moment van schrijven; als u het probeert kan dat anders zijn). Figuur 1.2 toont een deel van de resultaten. We zien dat de eerste links naar de website [www.ou.nl](http://www.ou.nl) van de Open Universiteit verwijzen.



FIGUUR 1.2 Deel van pagina met zoekresultaten

Heel gewoon, maar als we erbij stilstaan, toch ook verbazingwekkend. Hoe doet Google dat, 7 600 000 documenten vinden verspreid over het hele internet en dat in slechts 0,10 seconde? Een duizelingwekkende prestatie. We gaan in deze paragraaf globaal bekijken hoe Google werkt. Voor we daar aan toekomen, moeten we eerst kijken naar de ruimte die Google doorzoekt: het world wide web (www).

## 1.1 WEBPAGINA'S

We mogen ervan uitgaan dat iedereen die aan deze cursus begint, wel eens informatie 'van het internet' haalt, al dan niet met behulp van Google. En hoewel een zeer naïeve gebruiker zou kunnen denken dat al die informatie zich al in zijn computer bevindt, nemen we ook aan dat u weet dat dat niet zo is: de pagina's die u te zien krijgt, komen van buiten, 'van het internet'. Maar wat betekent dit nu eigenlijk? Daarmee bent u mogelijk minder bekend.

Later in deze cursus zult u nog veel meer leren over internet. Op dit moment is het voldoende om te weten dat het internet een netwerk is dat alle aangesloten computers met elkaar verbindt. Zodra u zich begeeft op het internet, maakt uw computer daar deel van uit.

De verbindingen tussen de computers worden voor verschillende toepassingen gebruikt. Bekende toepassingen zijn e-mail, het world wide web en chatten. Met behulp van e-mail sturen internetgebruikers berichten naar elkaar die later gelezen worden. Het world wide web ('het web') wordt onder andere gebruikt om informatie op te zoeken, spelletjes te spelen of inkopen te doen. Met behulp van chatprogramma's als MSN staan mensen in direct contact met elkaar en kunnen ze tekstberichten uitwisselen. Het internet is dus niet hetzelfde als het web: het is slechts één van de gebruiksmogelijkheden van het internet.

De basis van het gebruik van het web is: pagina's opvragen en bekijken. Die pagina's moeten wel eerst gemaakt zijn. Er zijn twee duidelijk verschillende rollen op het web: die van maker (producent) en die van gebruiker (consument) van pagina's.

Makers plaatsen hun pagina's onder een unieke naam op een computer waar ze opgehaald kunnen worden. Zulke computers worden webservers of kortweg servers genoemd. Als u in uw webbrowser de naam van zo'n pagina intypt of op een link naar die pagina klikt, dan wordt via allerlei tussenstappen de server gevonden waar die pagina op staat. Die verstuurt vervolgens via het internet die pagina naar uw computer, of beter gezegd naar uw webbrowser, en die vertoont de pagina op uw scherm. Om computers te vinden, wordt gebruik gemaakt van numerieke adressen; zowel de servers als uw eigen computer hebben zo'n adres. Figuur 1.3 illustreert de gang van zaken bij het ophalen van de hoofdpagina van Google.

Dit is een eenvoudig voorbeeld: de gezochte hoofdpagina is al kant-en-klaar aanwezig op de servers van Google. Het kan ook ingewikkelder. We typen in Google een zoekvraag ('John Martinkus'), klikken op zoeken en krijgen de pagina met resultaten. Deze is niet, zoals de hoofdpagina, al bij voorbaat kant-en-klaar opgeslagen in één van de servers van Google. Waar komt die dan vandaan? We kijken naar het adresveld, dat laat zien wat we precies aan Google vroegen. Daar staat (onder andere):

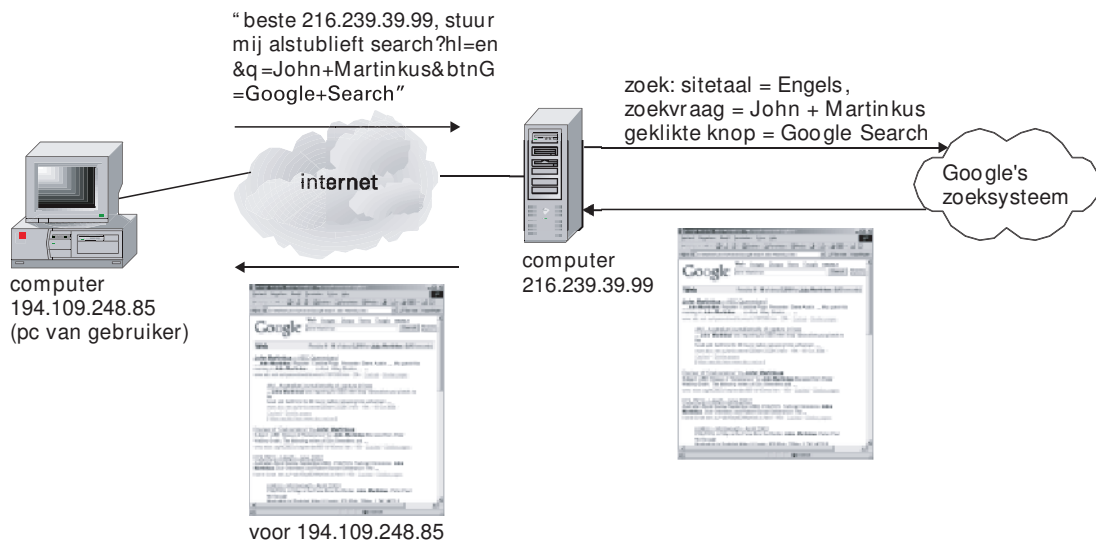
<http://www.google.nl/search?hl=en&q=John+Martinkus>



FIGUUR 1.3 Een pc haalt de hoofdpagina van Google op

Het eerste stuk, `www.google.nl`, verwijst naar de webserver van Google; dit wordt in figuur 1.3 vertaald in het numerieke adres 74.125.79.99. Achter de schuine streep staat eerst 'search'. De server weet daardoor dat het om een standaard zoekopdracht gaat. Na het vraagteken volgen de details van de zoekopdracht. De taal is Engels (`hl = en`) en de zoekvraag is `John+Martinkus` (`q=John+Martinkus`).

De computer van Google beantwoordt de zoekvraag niet zelf, maar geeft deze door aan een achterliggend systeem. Dat gaat aan de slag (we gaan straks bekijken hoe) en levert het resultaat aan in de vorm van een nieuwe webpagina, die vervolgens als antwoord over het internet naar uw computer wordt gestuurd. Figuur 1.4 illustreert het proces.



FIGUUR 1.4 Het antwoord op een zoekvraag wordt door een achterliggend systeem bepaald

## OPGAVE 1.1

Ga naar de Nederlandse Google-site ([www.google.nl](http://www.google.nl)), klik op Afbeeldingen en zoek naar plaatjes van flamenco (de bekende combinatie van muziek, zang en dans uit Zuid-Spanje). Wat staat er nu in het adresveld van uw webbrowswer?

De hier geschetste gang van zaken is niet uniek voor Google. Als u de ANWB-routeplanner of de OV-reisinformatie raadpleegt, gaat het in grote lijnen net zo: het verzoek wordt opgevangen door een webserver, die een achterliggend systeem aan het werk zet dat een antwoord produceert in de vorm van een webpagina.

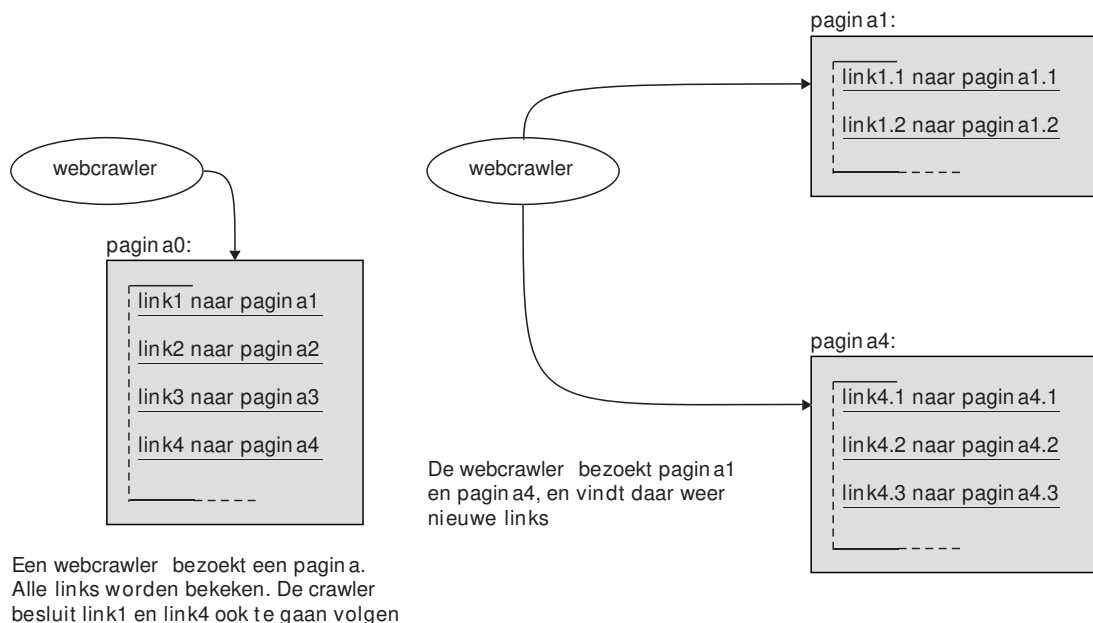
In de rest van deze paragraaf kijken we specifiek naar Google: wat zit er in die wolk rechts in figuur 1.4? In hele grote lijnen: Google verzamelt voortdurend informatie en slaat de resultaten daarvan op in een soort superwoordenboek, de index geheten, gesorteerd op relevantie. Het initiële succes van Google berust op dat laatste: Google deed dat sorteren op het moment van verschijnen veel beter dan de concurrentie. Het meeste werk is dus al gedaan voordat u een zoekvraag stelt. In de volgende deelparagrafen kijken we iets gedetailleerder naar hoe Google pagina's verzamelt en de index maakt.

We maken daarbij enig voorbehoud. Google is ontworpen door Larry Page en Sergey Brin, destijds studenten aan de Universiteit van Stanford. Zij publiceerden over de allereerste versie van hun zoekmachine, en een deel van de informatie in de volgende paragrafen is uit die publicaties afkomstig. In 1998 werd Google een bedrijf. De details van Google's werkwijze zijn sindsdien niet langer openbaar en zullen inmiddels zeker veranderd zijn. We beperken ons in deze leereenheid tot enkele basisconcepten van Google's zoekmachine, waarvan we redelijkerwijs kunnen aannemen dat ze nog steeds geldig zijn. De naam Google is afgeleid van het woord 'googol' dat het getal  $10^{100}$  aanduidt. De naam verwijst naar de immense hoeveelheid informatie die er op het web te vinden is. Bij de start van Google in 1998 bevatte de index 26 miljoen pagina's. Dat aantal is in de loop der jaren enorm gegroeid. In november 2004 bevatte de index 8 miljard pagina's en in juli 2008 werd de grens gepasseerd van 1 biljoen ( $10^{12}$ ).

## 1.2 WEBCRAWLERS

Laten we eerst eens kijken hoe Google webpagina's verzamelt. Het woord web in 'world wide web' verwijst niet naar het netwerk van computers die tot het internet behoren, maar naar het netwerk van pagina's. De meeste webpagina's bevatten hyperlinks (kortweg 'links' genoemd) naar andere pagina's, die op hun beurt weer naar andere pagina's verwijzen, enzovoort. Alle pagina's op alle webservern vormen dus een gigantisch web, waarvan de draden gevormd worden door de links tussen de pagina's.

Webcrawlers (ook wel spiders of webscrapers genoemd) volgen die draden om zoveel mogelijk webpagina's te bezoeken. De naam webcrawler lijkt te verwijzen naar iets dat daadwerkelijk van de ene computer naar de andere over het internet zwerft, zoals de virussen die zich via e-mail verspreiden. Dat is echter niet zo. Een webcrawler draait op één computer en lijkt op een browser. Anders dan in een browser gebeurt het ophalen van pagina's echter niet op verzoek van een menselijke gebruiker, maar automatisch. Een webcrawler vindt steeds nieuwe pagina's door de links te volgen op pagina's die al gevonden waren. Op den duur komt de crawler daardoor op grote delen van het world wide web. Figuur 1.5 illustreert het proces.



FIGUUR 1.5 Een webcrawler haalt steeds nieuwe pagina's op door links te volgen op eerder gevonden pagina's

#### OPGAVE 1.2

Kunt u bedenken waarom de webcrawler niet eenvoudigweg alle links op een pagina volgt?

Een webcrawler ontwerpen is niet zo eenvoudig als figuur 1.5 suggereert. We noemen enkele voor de hand liggende problemen.

- Er moet vermeden worden dat servers overbelast raken door de activiteiten van de webcrawler (zie de uitwerking van opgave 1.2).
- De webcrawler volgt maar één link tegelijk en moet dus de links onthouden die nog aan de beurt moeten komen. In figuur 1.5 bijvoorbeeld, zou de webcrawler eerst link1 kunnen volgen en link4 onthouden voor later. Bij pagina1 aangekomen, besluit de webcrawler dat link1.1 en link1.2 beide gevolgd moeten worden. Er zijn dan drie links die nog aan de beurt moeten komen, namelijk link4, link1.1 en link1.2. Steeds wordt één link uit die lijst gekozen en worden er na het bezoek aan de gekozen pagina nieuwe links toegevoegd. Het keuzecriterium is een belangrijk aspect van het

ontwerp van de webcrawler. Om de lijst niet tot in het oneindige te laten groeien, moet er ook een stopcriterium zijn, bijvoorbeeld om geen pagina's op te halen die verder dan een bepaald aantal stappen verwijderd zijn van het begin van de zoektocht.

- Komt de webcrawler op den duur wel overal, of worden grote delen van het web gemist?

Voor de eerste twee problemen heeft Google oplossingen gevonden in hun webcrawler (of eigenlijk webcrawlers, want ze gebruiker er meer dan één). Het derde probleem is een korte beschouwing waard.

Volgens een populaire theorie (de 'small-worldtheorie') is iedereen gemiddeld slecht zes stappen verwijderd van elke andere aardbewoner via een keten van mensen die elkaar ooit een hand hebben gegeven. Voor het web bestaat ook zo'n schatting: iedere webpagina is ongeveer twintig klikken verwijderd van iedere andere, en dat getal groeit slechts langzaam. Dat suggereert dus dat de meeste pagina's wel gevonden worden. Echter: dit geldt niet voor pagina's waarnaar geen directe links wijzen. Voorbeelden zijn pagina's die pas worden aangemaakt als iemand er om vraagt, zoals de informatie over een bepaalde route in de ANWB-routeplanner, of pagina's in een afgeschermd omgeving van een bedrijf. De webcrawler van Google mist zulke pagina's. Omdat het op deze manier ontsluiten van informatie via het web in belang toeneemt, is dit een serieus probleem.

### 1.3 HET MAKEN VAN DE INDEX

De pagina's die worden gevonden door de webcrawlers, worden door Google eerst opgeborgen in een lokaal depot en dan op verschillende manieren verder bewerkt. Ten eerste krijgt elke pagina een uniek nummer, de document-id. Ten tweede wordt de pagina (of een flink deel daarvan) ontleed. Alle woorden die erin voorkomen, worden opgeslagen onder de document-id, waarbij informatie wordt opgenomen over de plaats in het document en over het belang van het woord (bijvoorbeeld of het een tussenkop is, of vet staat). Daarmee ontstaat dus een afbeelding van pagina's op woorden. Vervolgens wordt die afbeelding geïnverteerd (omgedraaid) naar wat uiteindelijk nodig is: een afbeelding van woorden op pagina's (zie figuur 1.6).

d3797	Amsterdam	Amsterdam	d3797	d5988	...
	rondvaart	opera	d5988	...	...
	VVV	parkeren	d5988	...	...
	....	rondvaart	d3797		
		stadhuis	d5988	...	...
d5988	Amsterdam	VVV	d3797	...	...
	opera	...	...	...	...
	parkeren	...	...	...	...
	stadhuis				
	...				

FIGUUR 1.6 Voorwaartse en geïnverteerde index



Figuur 1.6 illustreert de twee afbeeldingen in sterk vereenvoudigde vorm (er is afgezien van alle extra informatie die Google over de woorden bijhoudt). Links is de afbeelding van documenten op woorden getoond. Uit deze afbeelding blijkt bijvoorbeeld dat document d3797 de woorden Amsterdam, rondvaart en vvv bevat, en document d5988 de woorden Amsterdam, opera, parkeren en stadhuis. Door die afbeelding om te draaien (figuur 1.6 rechts) kunnen we meteen zien in welke documenten het woord Amsterdam voorkomt. Bij het afhandelen van een zoekvraag wordt zo'n geïnverteerde vorm gebruikt.

Ten derde worden alle links in de pagina opgespoord. Deze zijn nodig voor de webcrawler om nieuwe pagina's te vinden, zoals we in paragraaf 1.2 hebben gezien. De links worden echter ook nog op andere manieren gebruikt. Kijk, ter illustratie, naar figuur 1.7. Deze toont een van de resultaten als we in Google zoeken op het onderwerp Amsterdam. De link wijst naar de website [www.uva.nl](http://www.uva.nl). De link zelf is niet betekenisvol voor mensen die de afkorting uva niet kennen. Uit de omschrijving die Google bij de link toont, blijkt dat het gaat om de website van de Universiteit van Amsterdam, en dat we op die website informatie kunnen vinden over het opleidingsaanbod. Deze omschrijving is afkomstig van de pagina waarop de link staat, en niet van de pagina [www.uva.nl](http://www.uva.nl) zelf. De tekst bij een link op een pagina wordt dus in de index opgeslagen bij de pagina waarnaar verwezen wordt. Het opslaan van omschrijvingen bij de links in de index, is één van de slimme ideeën die aan Google ten grondslag liggen.

### University of Amsterdam: Home

[www.uva.nl/](http://www.uva.nl/) ▼

The UvA is the largest university in the Netherlands with over 30000 students divided over seven major faculties.

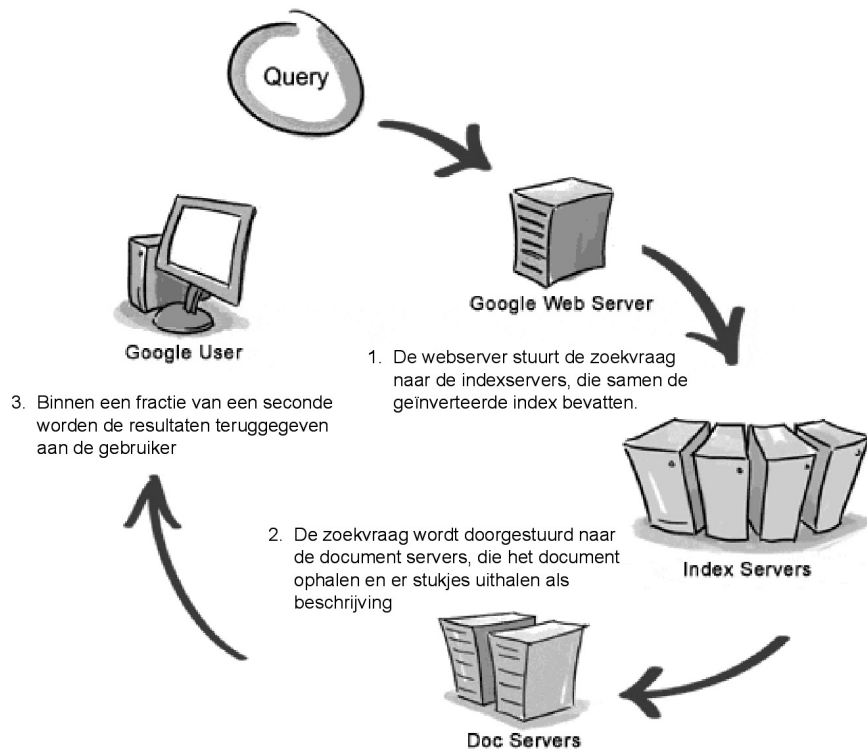
FIGUUR 1.7 De tekst van een link zegt veel over de pagina waarnaar verwezen wordt

De links op een pagina spelen ook een hoofdrol bij de waardering van zoekresultaten; dat was het belangrijkste nieuwe idee in Google. In de index staan bijvoorbeeld 88 500 000 documenten waarin het woord Amsterdam voorkomt, maar welke zijn nu de meest interessante en willen we dus als eerste zien in de lijst van zoekresultaten? In het begin slaagden zoekmachines er absoluut niet in, om een keuze te maken waarmee gebruikers tevreden waren. Google was de eerste zoekmachine die niet afging op de sleutelwoorden die de makers van de pagina opgeven, of op het aantal keren dat een woord in de pagina voorkomt. In plaats daarvan wordt aan iedere pagina een waardering toegekend (de zogeheten PageRank), die gebaseerd is op het aantal malen dat er vanuit een andere pagina naar verwezen wordt. Hoe meer links naar de pagina, hoe hoger de waardering. Niet iedere link telt daarbij echter even zwaar: links van een pagina die zelf een hoge waardering heeft, tellen zwaarder. Een link vanaf de startpagina van The New York Times telt daardoor bijvoorbeeld veel zwaarder dan een link vanuit de eerste eigen website van een scholier (omdat veel meer websites zullen linken naar The New York Times dan naar die

scholierenpagina). Ook tellen links van buiten een website zwaarder dan links binnen de website zelf. Dat maakt de berekening van de pagerank wel ingewikkeld. Want hoe weten we dat de startpagina van The New York Times een hoge waardering heeft, als we alle waarderingen juist aan het berekenen zijn? De waardering van iedere pagina heeft invloed op de waardering van iedere andere pagina, dus waar moeten we beginnen? Toch blijkt er wel degelijk een goede berekening mogelijk. De resultaten van pagerank waren spectaculair: opeens kregen gebruikers wel te zien wat ze wilden zien. Niet altijd (zie ook paragraaf 1.6), maar wel veel vaker dan voorheen.

#### 1.4 HET AFHANDELEN VAN DE ZOEKVRAAG

We kijken tot slot nogmaals naar de afhandeling van een zoekvraag, waarbij we nu een plaatje gebruiken dat afkomstig is van de website van Google zelf (zie figuur 1.8); de teksten zijn door ons vertaald.



FIGUUR 1.8 Afhandeling van een zoekvraag

Zoekvragen zijn niet altijd enkele woorden. Wanneer meer dan een woord wordt opgegeven, bijvoorbeeld 'open universiteit', wordt gezocht naar pagina's waar beide woorden op voorkomen, waarbij pagina's die ze ook nog na elkaar in die volgorde bevatten, een streepje voor krijgen. De resultaten uit de index moeten dus nog gecombineerd worden. Ook een zoekvraag als 'universiteit -open' is mogelijk; die levert pagina's op waarop wel het woord 'universiteit', maar niet het woord 'open' voorkomt.

Het onderste gedeelte van figuur 1.8 met de indexservers en de documentservers (docservers), komt overeen met de wolk 'Google's zoekstelsel' uit figuur 1.4. Elk van deze servers is een aparte computer. Ze heten servers omdat ze, net als de web servers, een voortdurende stroom verzoeken afhandelen. Figuur 1.8 toont vier indexservers en twee docservers, maar in werkelijkheid gaat dat om heel andere aantallen.

De precieze aantallen worden door Google niet bekend gemaakt. Bronnen op internet maken melding van ruim 30 rekencentra verspreid over de gehele wereld, onder andere in de Eemshaven in Nederland en in Mons (Bergen) in België, met in totaal meer dan 900 000 servers in 2011.

#### 1.5 EFFICIËNT OMSPRINGEN MET RUIMTE EN TIJD

We hebben slechts een globaal beeld gegeven van hoe Google werkt en veel details laten we noodgedwongen achterwege. Het wonder is niet zozeer dat het werkt (al is ook daar veel vernuft aan te pas gekomen), maar vooral dat het zo snel werkt, ondanks het feit dat het om onvoorstelbare hoeveelheden informatie gaat. De index van Google bevat meer dan een biljoen pagina's en er worden per dag honderden miljoenen zoekvragen afgewerkt. Alles bij het ontwerpen van zo'n zoekstelsel draait dus uiteindelijk om efficiënt omspringen met ruimte en met tijd. We noemen een paar vragen.

- Hoe kan men een webpagina in zo min mogelijk ruimte opbergen? Pagina's worden daarbij gecomprimeerd, samengeperst zonder verlies van informatie. Hoe gaat dat, en vooral: hoe gaat het snel?
- Het ophalen van een pagina van het web kost veel meer tijd dan hem uit het lokale depot (docserver) halen. Het kan dus de moeite waard zijn een pagina een tijdje in dat depot te houden en hem daar uit te halen als hij nodig is. Welke pagina's moeten worden vastgehouden? Hoelang? Is het acceptabel dat de echte pagina intussen is veranderd?
- Het ophalen van informatie van een lokale harde schijf is weer veel duurder dan iets uit het centrale geheugen van de computer ophalen. Als informatie van de harde schijf dus meer dan eens nodig is, dan moet ervoor gezorgd worden dat die tussendoor in het geheugen blijft. Hoe is dat te organiseren?
- Hoe moeten de voorwaartse en de geïnverteerde index georganiseerd zijn om zo min mogelijk ruimte te gebruiken? En hoe om er zo snel mogelijk iets in te kunnen vinden? En hoe zijn die twee tegen elkaar af te wegen?
- Hoe moeten de rekencentra, verspreid over de wereld, worden ingericht? En hoe kan het energieverbruik worden beperkt? Google let tegenwoordig bij het kiezen van locaties voor nieuwe rekencentra vooral op de vraag of in de enorme energiebehoefte kan worden voorzien.

We gaan hier zelfs geen poging doen om die vragen te beantwoorden. Als u meer wilt weten, kunt u de bronnen raadplegen die vermeld staan op de cursussite; het technisch gehalte daarvan is in het algemeen hoog.

## 1.6 GOOGLE EN DE WERELD

Tot nu hebben we Google beschreven als een interessant technisch hoogstandje, en niet ten onrechte: dat is het ook. Maar Google opereert natuurlijk niet in een vacuüm. We noemden in de introductie al de invloed op de taal: zowel in het Engels als in het Nederlands is googelen een werkwoord aan het worden. Het gebruik van Google is daarbij niet eens meer essentieel; iemand googelen betekent gewoon informatie over diegene opzoeken op het web. Maar er staat meer op het spel. Zoekmachines zoals Google hebben macht. Wij zoeken het web af met Google, en wat Google niet vindt, dat zien we niet. Voor iedere website die het moet hebben van veel nieuwe bezoekers, is het van buitengewoon groot belang om door Google gevonden te worden én om een zo hoog mogelijke paginawaardering te krijgen. Dus doet iedere websitemaker zijn uiterste best om Google te behagen. Google sleutelt nog regelmatig aan de manier waarop pagina's gewaardeerd worden; dat betekent dat soms een site die tijden (bij een bepaalde zoekterm) hoog bij de eerste tien stond, opeens verdwijnt. Updates van het Google zoekstelsel dragen namen; vooral de Florida-update van november 2003 is berucht. De wijzigingen in de waardering leidden tot woedende reacties en complottheorieën ('ze doen het expres'), temeer daar de details van de paginawaardering en van de crawlstrategie inmiddels (uiteraard) bedrijfsgeheimen zijn.

Manipuleert Google de resultaten? Daar zijn geen harde aanwijzingen voor, maar wel heeft het bedrijf grote commerciële belangen. Google is een beursgenoteerd bedrijf en heeft als zodanig een volstrekt andere positie dan het groepje onderzoekers dat het prototype midden jaren negentig ontwikkelde aan de universiteit van Stanford. Veel van het wantrouwen tegen Google richt zich op een mogelijke bevoordeling van de commerciële sector. Google doet gelukkig niet aan pop-upadvertenties. Wel tonen zoekvragen op Google inmiddels zogenaamde gesponsorde links, betaalde advertenties gekoppeld aan de zoekterm. Dit vormt een grote inkomstenbron voor Google.

Ook wordt er getwijfeld of Google bepaalde informatie achterhoudt of filtert. Dat speelt niet alleen in minder democratische landen, waarin overheden de informatie op het world wide web censureren en ongewenste sites blokkeren, maar eveneens in de westerse wereld. De vraag is of manipulatie wel per definitie verwerpelijk is. Willen we dat er vrijelijk naar kinderporno, zelfmoordmethoden en recepten voor explosieven kan worden gezocht, of hebben zoekmachines hier een taak? En zo ja, aan welke regels moeten ze dan voldoen en waar ligt de grens?

Privacy is daarnaast een belangrijke vraag. Google is, naast haar zoekmachine, ook één van de grootste advertentie-tussenpersonen op het internet. Google verdient zijn geld hoofdzakelijk met internetadvertenties en richt zich dus op het verzorgen van advertenties die zo dicht mogelijk zitten bij wat een gebruiker zoekt. Daarvoor gebruikt Google informatie over gebruikers, die ze van veel plaatsen kunnen verzamelen. Zo biedt Google gratis analyse-software voor websites aan, die beheerders makkelijk in staat stelt om te kijken hoe mensen hun website beleven. Maar: dat kan alleen maar omdat al die gegevens doorgespeeld worden aan Google. Daarnaast hebben veel websites filmpjes van YouTube opgenomen – maar ook YouTube is van Google. Al met al: als u een willekeurige webpagina bezoekt, is de kans groot dat Google dat te horen krijgt. Wat dat betekent voor uw privacy is helaas onduidelijk.

De conclusie van dit alles is, dat het niet alleen technische vragen zijn die we in deze paragraaf wel stellen maar niet beantwoorden: er zijn evenzeer juridische en ethische vragen te stellen. We hebben hier Google als een concreet voorbeeld genomen, maar zulke vragen zijn natuurlijk ook van toepassing op andere zoekmachines en internetbedrijven.

## 2 Het vak informatica

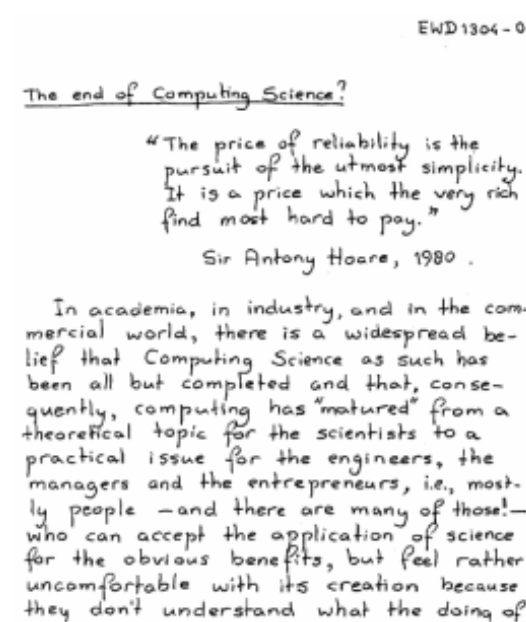
Google is een regelrecht product van het vakgebied informatica. In de vorige paragraaf zijn we al een aantal onderwerpen tegengekomen die allemaal met informatica te maken hebben. In deze paragraaf onderzoeken we nader wat het vakgebied informatica inhoudt. Eerst staan we stil bij de vraag waar het in het vakgebied nu eigenlijk om gaat (paragraaf 2.1) en vervolgens maken we een indeling in deelgebieden (paragraaf 2.2).

### 2.1 WAT IS INFORMATICA?

Informatica is een relatief jong vakgebied; beoefenaren zijn het nog verre van eens over de inhoud. Is informatica als wetenschappelijke discipline een exacte wetenschap, die theorieën formuleert en toetst over haar studieobject? En zo ja, wat is dat studieobject precies? Is het een ingenieursdiscipline, die methoden ontwikkelt om dingen (computers, maar ook programma's) te maken en hun geschiktheid te toetsen? Heeft het beide aspecten, die zich dan tot elkaar verhouden als bijvoorbeeld aerodynamica en vliegtuigbouwkunde? Of zijn deze kanten van ondergeschikt belang en moet het gebruik dat mensen maken van informaticaproducten centraal gesteld worden? Laten we eens naar twee uitersten in dit spectrum kijken.

Edsger W. Dijkstra (1930-2002) was een van de pioniers van het vakgebied en de enige Nederlander die de hoogste informaticaprijs, de Turing award, heeft gewonnen (in 1972). Hij had zeer uitgesproken opvattingen over wat het vakgebied wel en niet inhield. In zijn ogen was informatica een onderdeel van de wiskunde, en wel het onderdeel dat zich bezighoudt met de beheersing van complexiteit. Van hem stamt het volgende aforisme: 'De computerwetenschap gaat net zo

min over computers als de astronomie over telescopen.’ Met andere woorden: computers zijn hulpmiddelen bij het onderzoeken waar het in de informatica echt om gaat, maar niet het studieobject. Programmeren zag hij als een wiskundige discipline, waarbij de eigenschappen van het programma in de vorm van wiskundige stellingen bewezen moesten worden. Programma’s op een computer uitproberen om de fouten eruit te halen (testen) was hem dan ook een gruwel: vooraf moet immers al bewezen zijn dat het programma deugt. Informatica hield zich bovendien niet bezig met welk toepassingsgebied dan ook; zo iets als medische of juridische informatica was in zijn opvatting net zo onbestaanbaar als een vierkante cirkel.



FIGUUR 1.9 Meer dan in formele publicaties legde Dijkstra zijn inzichten neer in meer dan 1300 ‘EWD’tjes’, handgeschreven notities die via een informeel circuit gekopieerd en verspreid werden.

We zullen niet in detail op Dijkstra’s rechtvaardiging van zijn opvatting ingaan, maar het is wel goed als u zich realiseert dat bepaalde fundamentele resultaten uit de informatica dateren van voor de uitvinding van de computer, en dat ze die uitvinding ook mede mogelijk hebben gemaakt. Al in de jaren dertig zijn belangrijke eigenschappen van automaten (het woord computer verwees toen nog naar mensen) wiskundig beschreven en bleken er harde theoretische grenzen te zijn aan wat een automaat wel en niet kan. Dat informatica over computers gaat, is dus niet zo vanzelfsprekend als het lijkt.

Anderzijds is de rol van computers toch wel heel anders dan die van telescopen. Telescopen worden geconstrueerd met slechts één doel: de astronomie ondersteunen. Computers zijn niet gemaakt om een deelgebied van de wiskunde te ondersteunen en zijn bovendien alomtegenwoordig.

Wat in verband daarmee opvalt, is dat in Dijkstra's visie op informatica niet alleen de computers ontbreken, maar ook de mensen. Managers die inzichten over de betekenis van automatisering voor de processen in hun bedrijf willen, toekomstige gebruikers die programma's graag zo ontwikkeld willen zien dat ze er goed mee uit de voeten kunnen, spelers van spelletjes die hopen op verbetering van grafische technieken, ontwerpers van zoekmachines die worstelen met vragen rond privacy: zij allen moeten niet bij Dijkstra zijn.

Aan de andere kant van het spectrum staan opvattingen die mensen centraal willen stellen in de informatica. Zij beargumenteren dat het in de informatica bij uitstek over interactie zou moeten gaan. Een voorbeeld is Terry Winograd, hoogleraar in Stanford en werkzaam in de informatica sinds de jaren zestig toen hij bekendheid verkreeg met een programma dat (zeer beperkte) opdrachten gegeven in gewoon Engels kon uitvoeren. Zijn belangrijkste huidige interesse is het ontwerpen van de communicatie tussen mens en computer; als zodanig was hij zijdelings betrokken bij de ontwikkeling van Google en is hij nog steeds technisch adviseur bij dit bedrijf. In 1997 schreef hij, in een artikel over toekomstige trends in de informatica, onder meer:

'With the recent – and quite sudden – emergence of mass-appeal internet-centered applications, it has become glaringly obvious that the computer is not a machine whose main purpose is to get a computing task done. The computer [...] is a machine that provides new ways for people to communicate with other people.'

— Terry Winograd

In feite, gaat hij verder, was deze trend al veel eerder duidelijk. Het internet werd ontwikkeld om berekeningen over ver uit elkaar gelegen computers te kunnen spreiden, met als belangrijkste doel dat essentiële systemen niet met een aanval op één plek uitgeschakeld zouden kunnen worden (dit speelt tijdens de koude oorlog). Zodra het internet er was, werd het binnen de kortste keren echter vooral gebruikt om te e-mailen (e-mail is veel ouder dan het web, al werd het aanvankelijk niet door particulieren gebruikt). Ook de pc werd van het begin af aan vooral gebruikt om communicatie te ondersteunen; denk bijvoorbeeld aan tekstverwerkers. Steeds werd dus iets uitgedacht dat vervolgens min of meer onverwacht voor totaal andere, communicatieve doeleinden werd gebruikt. De vraag of de focus van de betreffende wetenschappelijke discipline dan soms verkeerd ligt, is daardoor alleszins gerechtvaardigd. Winograd ziet het verleggen daarvan overigens niet direct binnen de informatica gebeuren:

'The methods, skills, and techniques concerning these human aspects are generally foreign to those of mainstream computer science, and it is likely that they will detach (at least partially) from their historical roots to create a new field of 'interaction design'.'

— Terry Winograd

Zo'n visie leidt tot een ander soort onderzoek dan dat in wat Winograd de hoofdstroom noemt.

Deze twee uitersten (beide niet representatief) illustreren hoe ver de opvattingen over het vakgebied uiteen kunnen lopen. U moet zich daarbij wel realiseren dat het hier voor een deel een gebrek aan overeenstemming betreft over terminologie. Vrijwel niemand zal van mening zijn dat onderzoek naar nieuwe vormen van samenwerken op afstand of het exploreren van de ethische kanten van internetgebruik (zie paragraaf 1.6) niet op een universiteit thuishoort; waar men over van mening kan verschillen, is binnen welke discipline. En ook degenen die interactie centraal willen stellen, onderkennen het belang van zuiver technisch onderzoek. Wat zij constateren, is dat dat onderzoek tot nu toe zelden gedreven werd door wat uiteindelijk de echte behoefte bleek te zijn. Een interessante vraag daarbij is of aan dat laatste niet een wetmatigheid ten grondslag ligt die voortkomt uit inherente menselijke beperkingen. Niet alleen de bedenkers van internet, maar ook die van de drukpers, de stoommachine, de auto en de televisie voorzagen niet op welke wijze en hoe diepgaand hun uitvindingen de wereld zouden veranderen. Misschien is het eenvoudigweg onmogelijk om opinieweekbladen te bedenken zolang de drukpers niet bestaat, of om het world wide web te bedenken zolang er geen internet is.

Laten we gaan kijken naar de hoofdstroom van de informatica, of beter gezegd: de verschillende hoofdstromen. Als we kijken in de praktijk, dan treffen we ongeveer zes verschillende typen informatici aan (van Dael, 2001):

*Bestuurlijke informaticus*

– De *bestuurlijke informaticus* beschouwt het ontwerpen van informatiesystemen als de kern van de informatica. De bestuurlijke informaticus heeft kennis van de technische en de organisatorische kanten van het automatiseren van een bestuurlijk proces.

*Wiskundige informaticus*

– Voor de *wiskundige informaticus* is formeel-logische kennis van gegevensverwerking de kern van het vakgebied. Het gaat om toepassingsonafhankelijke kennis van gegevensverwerking met behulp van automaten. Wat geautomatiseerd wordt, maakt geen onderdeel uit van het vakgebied.

Er zijn overigens grote delen van de informatica die weliswaar onafhankelijk zijn van enig toepassingsgebied, maar niet of nauwelijks wiskundig van aard zijn en of om formele technieken vragen. We gebruiken daarom liever de term kerninformaticus voor de informaticus die zich richt op de toepassingsonafhankelijke kern van het vakgebied, waarbij wiskunde en formele technieken al dan niet een rol spelen.



<i>Technische informaticus</i>	– De <i>technische informaticus</i> combineert kennis van een te automatiseren technisch proces met kennis van de automatisering. Wat de bestuurlijke informaticus voor bestuurlijke processen doet, doet de technische informaticus voor de technische processen.
<i>Managende informaticus</i>	– Voor de <i>managende informaticus</i> is kennis van de invloed van (de organisatie van de) informatievoorziening op organisatieprocessen de kern van de informatica. Deze informaticus heeft slechts globale kennis van de technologie en heel veel kennis van de organisatieprocessen waarbinnen die technologie wordt toegepast.
<i>Beherende informaticus</i>	– De <i>beherende informaticus</i> staat voor de groep mensen die zorgen dat netwerken blijven draaien.
<i>Gebruikende informaticus</i>	– De <i>gebruikende informaticus</i> past de informatietechnologie op een geavanceerde wijze toe en benut die binnen een ander vakgebied.

## OPGAVE 1.3

Kunt u de informatici Dijkstra en Winograd ergens in deze opsomming plaatsen?

Onder de studierichtingen Technische informatica en Informatica aan de universiteiten bestaat een behoorlijke overeenstemming over de inhoud van de opleiding. Hoewel de definities op de websites van de universiteiten uiteenlopen, zullen de meeste zich kunnen vinden in een definitie als die van de International Federation of Information Processing, de IFIP.

*‘Informatics or Computer Science is a branch of knowledge or study, especially concerned with establishing and systematising facts, principles, and methods, as by experiments and hypotheses (...). Its field of study [is] the design and realisation of (programmable) systems. (...)’*

*Informatica is een tak van kennis of studie die zich bezighoudt met het vaststellen en systematiseren van feiten, principes en methoden, door middel van het uitvoeren van experimenten en het opstellen van hypothesen. (...). Het studieobject van de informatica is het ontwerp en de realisatie van (programmeerbare) systemen. (...)*

Wie informatica gaat studeren, wordt in de basis opgeleid tot kerninformaticus. Afhankelijk van de accenten die de opleiding legt, komt daarnaast ook de bestuurskundige en/of de technische variant aan bod. Beherende, managende en gebruikende informatici komen vooral van andere opleidingen (managementwetenschappen, economie of informaticaspecialisaties binnen het eigen vak).

In de volgende paragraaf bekijken we onderwerpen uit het vakgebied informatica zoals gedoceerd aan de universiteiten. U zult in die beschrijving zowel kenmerken zien van een theoretische discipline (waarin de nadruk ligt op onderzoeken) als van een ingenieursdiscipline (met de nadruk op construeren). De indeling is losjes gebaseerd op het al wat oudere Unified Classification Scheme for Informatics (UCSI). Ook al is het gedateerd, het biedt nog steeds een

goede basis voor uitleg. Dit schema kent drie indelingsniveaus: hoofddomeinen worden onderverdeeld in domeinen, en die weer in subdomeinen. Omdat we geen opsomming van termen willen geven die u veelal weinig zullen zeggen, nemen we alleen de hoofddomeinen letterlijk over. Hun beschrijving is echter wel geïnspireerd door de UCSI-indeling.

## 2.2 DEELGEBIEDEN VAN INFORMATICA

Als hoofddomeinen van de informatica benoemt UCSI: computersystemen, softwaresystemen, informatiesystemen, en context van informatica. We bespreken ze een voor een, en noemen steeds de belangrijkste deelgebieden binnen dat hoofddomein. Voor illustraties van het belang daarvan verwijzen we waar mogelijk naar Google.

### 2.2.1 *Computersystemen*

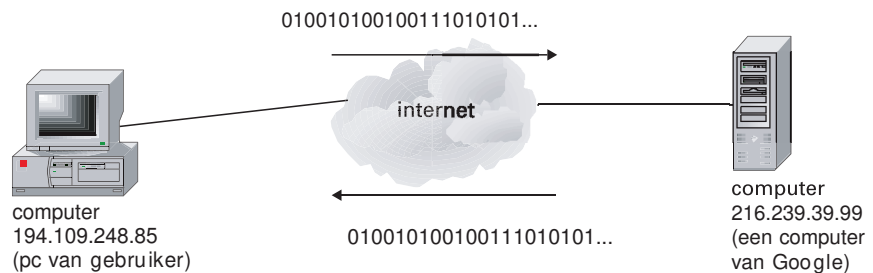
In het hoofddomein computersystemen ligt de nadruk op techniek. Het gaat in dat domein zo te zeggen om de harde basis van computers en andere 'intelligente' apparatuur. De technische informaticus zal veel kennis moeten hebben uit dit deelgebied, en ook de kerninformaticus heeft op dit terrein de nodige kennis en vaardigheden.

Een eerste onderwerp uit dit domein is het ontwerpen van digitale systemen (computers en andere intelligente apparaten). U moet daarbij niet in de eerste plaats aan de elektronische basiscomponenten zelf denken, maar meer aan de manier waarop hun samenwerking is georganiseerd, inclusief de communicatie met randapparatuur (printers, scanners, cd-romstations en ook harde schijven).

De relevantie van dit onderwerp voor het ontstaan van Google is in eerste instantie triviaal: zonder computers geen internet, zonder internet geen web en zonder web geen Google. In tweede instantie is er ook een meer specifieke invloed, die een gevolg is van het feit dat Google opereert aan de rand van de technische mogelijkheden. In paragraaf 1.5 gaven we al aan dat belangrijke problemen bij de constructie van Google vrijwel onbesproken zijn gebleven, en dat die te maken hebben met het efficiënt omspringen met ruimte en tijd. Op welk punt zulke problemen zich voordoen en of en hoe ze kunnen worden aangepakt, wordt volledig bepaald door de onderliggende technologie. Het is bijvoorbeeld niet ondenkbaar dat in de toekomst het onderscheid tussen centraal geheugen (in de computer) en geheugen op de harde schijf wegvalt. Daarmee wordt ook de vraag minder relevant hoe het opvragen van informatie van de harde schijf tot een minimum beperkt kan worden.

*Computernetwerken*

Ook het ontwerpen en laten functioneren van *computernetwerken* behoort tot dit domein. Laten we om dit domein te verduidelijken nog eens kijken naar figuur 1.3, maar nu in een iets andere vorm (zie figuur 1.10).



FIGUUR 1.10 Informatie wordt verstuurd over het internet

Zowel het verzoek aan Google als de teruggestuurde pagina worden verstuurd als pakketjes met nullen en enen. De pc en de server van Google zijn niet rechtstreeks met elkaar verbonden; daar zitten allerlei andere computers tussen. Een computer midden in het netwerk krijgt dus van allerlei kanten pakketjes met nullen en enen binnen en moet uitzoeken of die voor de computer zelf bestemd zijn of moeten worden doorgestuurd, en zo ja, welke kant op. Hoe bepalen we de weg door het netwerk die de pakketjes moeten afleggen? Soms is er ergens storing, zodat er pakketjes kwijtraken of verminkt worden doorgegeven. Hoe merkt de ontvanger dat en wat moet die dan doen? En als dat allemaal is opgelost en we dus foutloos pakketjes met nullen en enen van de ene computer naar de andere kunnen sturen, hoe maken we daar dan een toepassing als e-mail of het web mee?

### Besturingssystemen

Een derde onderwerp is de ontwikkeling van systeemprogrammatuur of *Besturingssystemen*. Het gaat daarbij om programma's waarvan de gemiddelde gebruiker zich nauwelijks realiseert dat ze bestaan (behalve als ze niet goed werken, wat helaas nogal eens gebeurt). Besturingssystemen als Windows of Linux zijn verzamelingen systeemprogramma's. We noemen twee voorbeelden van het vrijwel onzichtbare werk dat door die programma's gedaan wordt. Stel, u opent een Word-bestand via dubbelklikken met de muis. De enige informatie die de computer krijgt, is dat er op een bepaalde plek op het scherm geklikt is. Er zijn systeemprogramma's voor nodig om uit te zoeken binnen welk venster dat was en wat daar stond. Als uitgevonden is dat het om een bestand ging, moet dat bestand gelokaliseerd worden op de harde schijf. Word moet gestart worden en delen van het bestand moeten van de harde schijf naar het centrale geheugen getransporteerd worden. Ondertussen bent u misschien ook een groot bestand van het web aan het downloaden. Het lijkt dan dus of uw computer twee dingen tegelijk aan het doen is, terwijl de centrale processor in de meeste computers eigenlijk maar één taak tegelijk uit kan voeren. Ook dat wordt door systeemprogramma's geregeld. Besturingssystemen zijn niet uit de lucht komen vallen en zijn ook zeker niet door Microsoft uitgevonden; veel achterliggende ideeën zijn het resultaat geweest van (universitair) onderzoek. Voor Google is zowel het onderwerp netwerken als het onderwerp systeemprogrammatuur van groot belang.

We zagen in paragraaf 1.4 dat de rekenkracht van Google geleverd wordt door grote aantallen onderling verbonden servers, die samen een van de krachtigste supercomputers ter wereld vormen. Om dat allemaal goed te laten werken, is gespecialiseerde kennis over netwerken en systeemprogrammatuur vereist.

Als laatste noemen we een onderwerp dat de laatste jaren steeds belangrijker wordt, namelijk dat van de zogenaamde embedded systems. Tot twintig jaar geleden was het redelijk duidelijk wat een computer was en wat niet. De grens tussen computers en de ingebouwde gegevensverwerking in consumentenelektronica is in hoog tempo aan het vervagen. Palmtopcomputers, mobiele telefoons, muziekspelers, dvd-recorders, wasmachines, koffiezetapparaten en auto's bevatten allemaal centrale verwerkingseenheden en hebben hun eigen systeemprogramma's en toepassingsprogramma's.

### 2.2.2 Softwaresystemen

*Programmeren*

In het domein softwaresystemen staat de technische kant van het *Programmeren* centraal. Het is daarmee een belangrijk werkkterrein van de kerninformaticus.

Computers werken uiteindelijk met nullen en enen, maar iemand die een programma voor een computer schrijft, schrijft dat niet in die vorm op. In plaats daarvan zijn er allerlei programmeertalen ontwikkeld waarin een programmeur de instructies voor de computer kan uitdrukken. De computer zet die instructies vervolgens zelf om in de benodigde nullen en enen. De vraag hoe programma's en in het verlengde daarvan programmeertalen er uit moeten zien, is een belangrijk onderwerp uit de softwaretechnologie. In de loop der jaren zijn op universiteiten en onderzoeksinstituten van grote bedrijven honderden programmeertalen ontwikkeld, vanuit allerlei verschillende ideeën over programmeren. Niet elke taal is daarbij even geschikt voor ieder soort programma. De ontwikkelaars van Google bijvoorbeeld hebben niet alle benodigde programmatuur in dezelfde taal geschreven. Voor de webcrawler gebruikten ze de taal Python, een vrij wiskundig georiënteerde taal waarin de lijst van links die de webcrawler moet bijhouden en de manieren waarop daaruit gekozen wordt (zie paragraaf 1.2), mooi compact zijn uit te drukken. Voor het zoekstelsel zelf gebruikten ze echter de verwante talen C en C++. Programma's in die talen zijn vaak wat minder overzichtelijk, maar leiden bij uitvoering tot een hogere snelheid (wat voor het zoeken essentiëler is dan voor het crawlen).

*Algoritmen*

Binnen het domein van de softwaresystemen valt ook het onderzoek naar *algoritmen*, oplossingsmethoden voor bepaalde problemen (sorteren is een standaardvoorbeeld). Ook worden wiskundige technieken ontwikkeld om te kijken wat voor eigenschappen zo'n oplossing heeft (de ene sorteermethode is bijvoorbeeld sneller dan de andere). Het belang daarvan kunnen we gemakkelijk illustreren aan de hand van Google.

Een webcrawler is een programma dat verschillende paden volgt in een netwerk; aan dat probleem was al veel gedaan lang voor het web was uitgevonden. We zagen (opgave 1.2) dat een slecht georganiseerde webcrawler veel overlast kan veroorzaken op het internet; het is daarom van groot belang dat programmeurs van webcrawlers vooraf goed begrijpen hoe hun programma zich zal gedragen. Ook pagerank, de berekening van de paginawaardering, steunt op theorie. In paragraaf 1.3 vermeldden we al, dat de waardering van een bepaalde pagina in feite afhangt van de waardering van alle andere pagina's op het web. Kan die dan wel berekend worden? Door het probleem wiskundig te beschrijven, kan men inzien dat het mogelijk is om stap voor stap dichterbij een uiteindelijke waardering te komen.

Een programma om een printer aan te sturen, kan door één programmeur in korte tijd worden ontwikkeld. De programmeur moet er wel voor zorgen dat het programma goed kan samenwerken met andere systeemprogramma's, maar dat is allemaal te overzien. Bij de ontwikkeling van een volledig bedrijfssysteem of een systeem voor de patiëntenadministratie van een ziekenhuis spelen hele andere problemen. Zulke *systemen worden ontwikkeld* door tientallen programmeurs, die er een jaar of langer aan werken. Hun bruikbaarheid wordt niet alleen bepaald door de technische eigenschappen van het programma, maar net zo zeer door de mate waarin de beoogde gebruikers er mee uit de voeten kunnen. Zulke systemen moeten jaren kunnen meegaan en dus kunnen meegroeiën met veranderingen. Om zulke systemen te ontwikkelen, is een goed bouwplan nodig, met een goede rolverdeling tussen allerlei betrokkenen. Binnen het deelgebied software engineering wordt nagedacht over hoe dat het beste kan verlopen.

De theoretische informatica bestudeert grondslagen van het vakgebied. We vermeldden al dat belangrijke resultaten op dit gebied dateren uit de tijd dat computers nog niet bestonden, en informatica (al bestond die naam toen ook nog niet) echt nog een deelgebied van de wiskunde was. Er werd bijvoorbeeld aangetoond dat er voor bepaalde problemen helemaal geen systematische oplossing kan bestaan en dat er ook problemen bestaan die in theorie misschien wel systematisch oplosbaar zijn, maar in de praktijk altijd te veel rekentijd zullen vragen. Een voorbeeld van dit laatste is schaken. In theorie kan in iedere stand eenvoudig de beste zet worden gekozen door naar alle mogelijke voortzettingen te kijken. Helaas zijn dat er meer dan het geschatte aantal atomen in het heelal: deze strategie is praktisch onuitvoerbaar.

Een onderwerp dat wat moeilijk te plaatsen is, maar dat door UCSI tot het domein van de softwaresystemen gerekend wordt, is kunstmatige intelligentie. Over de definitie van dit deelgebied bestaat al net zoveel onenigheid als over de definitie van informatica als geheel, maar losweg gaat het hier om het construeren van programma's voor taken die slecht systematisch oplosbaar zijn. Dat geldt niet alleen voor een op zich goed gedefinieerde taak als schaken, maar ook voor allerlei taken die mensen uitstekend kunnen uitvoeren zonder precies te kunnen omschrijven wat de taak inhoudt en hoe die wordt aangepakt.

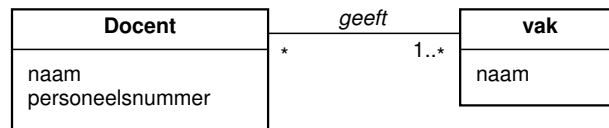
Voorbeelden zijn: door een kamer lopen zonder ergens tegenaan te botsen, Nederlands (of een andere taal) spreken en verstaan, autorijden, een handgeschreven brief lezen of achterhalen waarom een auto niet wil starten. In de jaren zestig waren onderzoekers zeer optimistisch over het vermogen van computers om ook dit soort taken uit te voeren. Men dacht toen bijvoorbeeld nog dat er voor het jaar 2000 een computer zou zijn die in een (getypte) conversatie niet van een mens te onderscheiden zou zijn. Dat is erg tegengevallen, maar technieken die binnen dit deelgebied zijn ontwikkeld, worden inmiddels wel op allerlei andere manieren gebruikt.

### 2.2.3 *Informatiesystemen*

Informatiesystemen vormen het derde hoofddomein van de informatica. In dit deelgebied gaat het om het construeren van geautomatiseerde systemen die processen in de buitenwereld ondersteunen. Verreweg de meeste daarvan zijn administratieve toepassingen: systemen die gebruikt worden door banken, in ziekenhuizen, in winkels, in garages en ga zo maar door. De bestuurlijke informaticus heeft bij uitstek verstand van de onderwerpen die binnen dit deelgebied vallen.

De meeste informatiesystemen zijn administratieve systemen, die werken met veel gegevens. Hier kan het systeem van Google als extreem voorbeeld dienen. De index van Google bevat informatie over zeer grote delen van het web. Die informatie wordt voortdurend aangepast en miljoenen keer per dag geraadpleegd. Hoe kunnen zulke grote hoeveelheden gegevens het best georganiseerd worden en hoe kan men wat gezocht wordt, er snel in vinden? En hoe wordt ervoor gezorgd dat het allemaal blijft kloppen, ook wanneer heel veel mensen of programma's tegelijkertijd toegang tot de gegevens willen hebben? Dit alles wordt onderzocht binnen het deelgebied dat zich bezig houdt met gegevensopslag, ofwel met databases.

De ontwerpers van Google hadden één groot voordeel bij het ontwerpen van hun gegevensverwerking: de gegevens waarmee Google werkt, spelen geen rol in bestaande bedrijfsprocessen. Voor Google is een dode link (u klikt op een zoekresultaat, maar die pagina blijkt er niet meer te zijn) of een gemiste pagina niet problematisch. In een ziekenhuis moeten echter altijd de volledige en juiste resultaten worden getoond bij het opvragen van patiëntengegevens. Ook mogen er geen problemen ontstaan wanneer patiëntengegevens onafhankelijk van elkaar op hetzelfde moment vanaf twee verschillende computers worden ingevoerd. In deze situaties moeten de processen en de gegevens die ze gebruiken, eerst zorgvuldig in kaart worden gebracht vóór ze opgenomen kunnen worden in een geautomatiseerd systeem. Dit in kaart brengen van informatie en processen, wordt modelleren genoemd. Het doel is om een duidelijke, ondubbelzinnige beschrijving te produceren (zie figuur 1.11) waarop het ontwerp gebaseerd kan worden van de database en van de programma's die samen het informatiesysteem vormen. Hiervoor zijn allerlei speciale methoden en notatietalen ontwikkeld.



FIGUUR 1.11 Voorbeeld van een (klein) informatiemodel.

Hieruit kan bijvoorbeeld worden afgelezen dat een docent meer dan een vak kan geven; zou dat niet kunnen, dan zou het model er anders uitzien.

Het laatste onderwerp dat we hier willen noemen en dat door UCSI gerekend wordt tot het domein van de informatiesystemen, is de interactie tussen mens en computer. De buitenkant van een programma of systeem, dat wat gebruikers te zien krijgen en waar ze mee moeten werken, wordt de gebruikersinterface genoemd. Hoe goed een programma is, hangt net zo sterk van die interface af, als van wat het programma allemaal kan. Een schoolvoorbeeld is pc's: pas toen ze bediend konden worden door met de muis op een grafisch scherm te klikken, konden ze doordringen in de huiskamers. Microsoft heeft zich pas relatief laat bij deze ontwikkeling aangesloten. Het idee is afkomstig van de universiteit van Stanford en van Xerox en vervolgens op grote schaal toegepast door Apple. Windows is daar regelrecht van afgekeken. Bij Google is veel aandacht besteed aan een goede interface. De startpagina is eenvoudig en de zoekresultaten worden overzichtelijk gepresenteerd. Voor de komst van Google konden bij andere zoekmachines de resultaten soms maar met moeite tussen de advertenties uitgevist worden. Google was ook de eerste zoekmachine die stukjes uit de gevonden pagina's liet zien met de zoektermen erop, zodat de gebruiker goed in kan schatten of een gevonden pagina de moeite van het bezoeken waard is.

Maar wat is precies een goede interface en hoe kunnen interfaces zo ontworpen worden dat ze aansluiten bij de manier waarop de beoogde gebruikers hun werkprocessen organiseren? Dat is een onderwerp waar nog veel onderzoek naar gedaan kan worden.

Ook de ondersteuning door computers van de interactie tussen mensen onderling vormt een interessant onderzoeksterrein, dat ook goed tot het domein Context van informatica gerekend kan worden (zie volgende paragraaf). Dit is het terrein waarop Winograd onderzoek doet en dat hij centraal zou willen stellen. Een onderwerp uit deze hoek dat momenteel sterk in de belangstelling staat, is computerondersteund samenwerken.

#### 2.2.4 Context van informatica

In paragraaf 1 over Google hebben we al gezien dat informaticaproducten niet louter technische systemen zijn, maar dat ze functioneren in een omgeving en daar ook op ingrijpen. De wisselwerking met hun omgeving vormt het onderwerp van het laatste hoofddomein. We noemen twee deelgebieden.

Een groot aantal onderwerpen uit dit deelgebied kan worden samengenomen onder de kop informatica en samenleving. Het gaat daarbij om economische, politieke, juridische, culturele en filosofische implicaties van het gebruik van computers en andere digitale systemen, waarbij er uiteraard een flinke overlap is met de aangrenzende wetenschapsgebieden. Paragraaf 1.6, over Google en de wereld, heeft hopelijk al voldoende duidelijk gemaakt wat het belang is van dit soort onderwerpen. Enige kennis uit dit gebied hoort tot de bagage van ieder type informaticus. Dit geldt bijvoorbeeld voor de beroepscode (wat geldt voor een informaticus als beroepsmatig ethisch handelen) en voor elementaire juridische aspecten, zoals de regelgeving rond privacy.

Automatisering leidt tot allerlei veranderingen in organisaties. Er verdwijnen functies en er komen nieuwe bij. Geautomatiseerde systemen moeten onderhouden worden. Er moeten nieuwe soorten beslissingen genomen worden: Kopen we een kant-en-klaar systeem of laten we het speciaal maken? Kopen we nieuwe computers, en zo ja, passen de oude softwaresystemen daar nog wel op? Wanneer moeten de medewerkers overschakelen op nieuwe software? In eerste instantie werden meestal alleen bestaande bedrijfsprocessen ondersteund met een geautomatiseerd systeem, maar later ontstonden er ook geheel nieuwe bedrijfsprocessen en zelfs bedrijven onder invloed van automatisering. Informatica en management houdt zich in brede zin bezig met wat dit voor bedrijven betekent. Dit is het terrein van de managende informaticus uit eerdergenoemde classificatie.

#### 2.2.5 Overige

Sommige onderwerpen uit het vakgebied informatica zijn moeilijk bij een van de vier hoofddomeinen onder te brengen, omdat ze raakvlakken hebben met (vrijwel) al deze domeinen.

Zo is bijvoorbeeld beveiliging bij uitstek interdisciplinair onderwerp, dat vooral sinds de opkomst van grootschalige netwerken (niet alleen internet, maar ook bijvoorbeeld het netwerk waarover alle pinbetalingen worden verstuurd) zeer belangrijk is geworden. Hoe zorgen we ervoor dat vertrouwelijke informatie die over een netwerk wordt verstuurd, ook vertrouwelijk blijft? Hoe weten we dat een bericht (bijvoorbeeld een betalingsopdracht) ook werkelijk afkomstig is van de afzender? Hoe houden we onze informatiesystemen gesloten voor onbevoegden en hoe regelen we de bevoegdheden van degenen die wel toegang hebben (een baliemedewerker van een ziekenhuis mag geen wijzigingen aanbrengen in de medische gegevens van de patiënten). Beveiliging speelt een rol bij het ontwerpen en inrichten van zowel netwerken als informatiesystemen. Beveiliging steunt onder meer op de beschikbaarheid van goede algoritmen voor het vercijferen (omzetten in geheimschrift) van informatie.



## 2.3 ETHIEK EN INFORMATICA

Computers raken meer en meer verweven met het dagelijks leven van mensen. Steeds meer zaken worden beheerd door een computer – of ze nu een toetsenbord en een beeldscherm hebben of niet. Zo bevat een moderne auto meer dan 30 microprocessors, met ieder hun eigen stukje software. Ook staan steeds meer van deze computers in verbinding met het internet. Televisies, modems, smartphones, printers, pinapparaten, babycamera's, smart watches, lampen, thermostaat. . . talloze stukjes software zitten ergens verstopt in een hoekje van de wereld, terwijl ze verbonden zijn aan internet. Vaak zitten er sensoren aan vast om de wereld waar te nemen, soms ook actuatoren om de wereld mee te beïnvloeden.

Kortom, software dringt door in alle hoeken en gaten van de 'echte' wereld en beïnvloedt tegelijkertijd steeds meer ons dagelijks handelen. Daarom is het belangrijk om je hiervan als informaticus bewust te zijn en ervoor te zorgen dat je ethisch handelt. De software die een informaticus ontwerpt, bouwt, of onderhoudt, bestaat waarschijnlijk uit allerlei stukken software van anderen. Daarnaast behandelt de software waarschijnlijk data die veel over mensen kan onthullen – hetzij direct (naam, adres, leeftijd. . .) hetzij indirect, doordat gedrag is af te leiden uit het gebruik van de software.

Vanuit de vakgroep Informatica van de Open Universiteit benadrukken wij hier dat een informaticus ethisch dient te handelen. Dit houdt in dat een informaticus overzicht moet houden over de impact van het door hem of haar uitgevoerde werk/onderzoek op anderen, en er zorg voor draagt dat het werk wordt uitgevoerd met inachtneming van de rechten van anderen. Dit geldt voor alle informatici: van software-ontwerper tot ethisch hacker.

De Amerikaanse beroepsorganisatie voor informatici, de Association for Computing Machinery (ACM), heeft een gedragscode opgesteld voor ethisch gedrag. Deze formuleert in een aantal kernpunten wat ethisch handelen betekent:

- Contribute to society and human well-being.
- Avoid harm to others.
- Be honest and trustworthy.
- Be fair and take action not to discriminate.
- Honor property rights including copyrights and patents.
- Give proper credit for intellectual property.
- Respect the privacy of others.
- Honor confidentiality.

— ACM Code of Ethics and Professional Conduct

Voor een uitgebreidere bespreking van deze punten kunt u terecht op de website van ACM, <http://www.acm.org>.

Ook tijdens de studie speelt ethisch gedrag een rol. Het kan bijvoorbeeld gebeuren dat studenten met de beste intenties plaatjes of een stukje tekst overnemen voor een werkstuk of rapport. Dat mag, zolang maar duidelijk is wat er precies is gekopieerd en er een

bronvermelding is opgenomen ('honor property rights' en 'give proper credit'). Dat geldt ook voor samenwerken: samen studeren is prima, de uitwerking van een andere student integraal overnemen uiteraard niet.

### 3 Het vak Inleiding Informatica

In deze inleidende informaticacursus kunnen niet alle onderwerpen aan bod komen die in paragraaf 2 genoemd zijn. Tabel 1.1 geeft een overzicht van de onderwerpen die wel worden behandeld, en laat zien waar deze passen in de indeling uit de vorige paragraaf. U zult zien dat er enkele onderwerpen zijn die in deze cursus toch verder niet ter sprake komen; dit geldt bijvoorbeeld voor embedded systems. Van de meeste andere onderwerpen komen steeds slechts enkele aspecten aan de orde. Dat is een bewuste keuze geweest; we hebben onderwerpen weggelaten om, enige diepgang aan te kunnen brengen in wat we wel behandelen.

TABEL 1.1 Welke onderwerpen komen in deze cursus aan bod?

<i>leereenheid</i>	<i>titel</i>	<i>hoofddomein</i>	<i>deelgebied(en)</i>
2	Systemen en systeemontwikkeling	informatiesystemen / softwaresystemen	modelleren; software engineering
3-4	Objectgeoriënteerd ontwikkelen: analyse en ontwerp	informatiesystemen	modelleren
5-8	Verzamelingenleer, bewerkingen op getallen, formeel beschrijven en een kennismaking met logica	softwaresystemen	theoretische informatica
9	Het besturingssysteem van een computer	computersystemen	systeemprogrammatuur
10	Imperatief programmeren	softwaresystemen	programmeren
11-12	Relationele databases	informatiesystemen	gegevensopslag
13-14	Cryptografie en Security	interdisciplinair	
15-16	Internet: toepassingen en structuur	computersystemen	computernetwerken

## S A M E N V A T T I N G

In grote lijnen is geschetst hoe één bepaalde informaticatoepassing werkt, namelijk de zoekmachine Google. Daarbij is ook aandacht besteed aan maatschappelijke aspecten van zo'n systeem.

We constateren dat er geen algemeen geaccepteerde definitie bestaat van het vakgebied informatica. Ter illustratie zijn twee uiterste standpunten genoemd: namelijk het standpunt van Dijkstra die informatica als een tak van de wiskunde zag, en het standpunt van Winograd die informatica het liefst zou willen opvatten als de wetenschap die zich bezighoudt met het ontwerpen van (vormen van) interactie tussen mensen. Uit onderzoek is gebleken dat er in Nederland zes typen informatici worden onderscheiden: de wiskundige of kerninformaticus, de bestuurskundige, de technische, de managende, de beherende en de gebruikende informaticus. We maakten een indeling van het vakgebied, gebaseerd op de wijze waarop informatica wordt onderwezen aan de meeste universiteiten. Daar worden veelal kerninformatici opgeleid met kennis van technische en/of bestuurlijke toepassingen. Deze indeling is niet uitputtend; niet alle deelgebieden zijn genoemd (maar wel de belangrijkste).

- Het hoofddomein computersystemen houdt zich bezig met de technische basis van computers. Deelgebieden zijn ontwerp van digitale systemen, computernetwerken, systeemprogrammatuur en embedded systems.
- Het hoofddomein softwaresystemen houdt zich bezig met de technische kant van het programmeren. Deelgebieden zijn programmeren en programmeertalen, algoritmen, software engineering, theoretische informatica en kunstmatige intelligentie.
- Het hoofddomein informatiesystemen houdt zich bezig met het construeren van geautomatiseerde systemen die processen in de buitenwereld ondersteunen. Deelgebieden zijn gegevensopslag, modelleren van informatie en processen, en interactie en presentatie.
- Het hoofddomein context van informatica houdt zich bezig met de wederzijdse beïnvloeding tussen informaticaproducten en de omgeving waarin ze functioneren. Deelgebieden zijn informatica en samenleving, en informatica en management.

Tot slot noemden we security – een onderwerp dat raakvlakken heeft met al deze domeinen.

Ten slotte zijn deze onderwerpen gerelateerd aan de rest van de cursus.



Z E L F T O E T S

- 1 Noem zes verschillende typen informatici en geef daarvan een korte karakterisering.
- 2 Geef voor elk van de volgende vier vragen aan, tot welk hoofddomein en welk deelgebied daarvan ze gerekend moet worden.
  - a Aan welke regels moeten bedrijven als Google zich volgens de wet houden bij het vergaren van informatie over de gebruikers van hun producten?
  - b Hoe wordt in kaart gebracht welke gegevens nodig zijn voor een bepaald bedrijfsproces (bijvoorbeeld het inschrijven van leerlingen op een school)?
  - c Hoe kan men een pc schijnbaar twee dingen tegelijk laten doen, zoals een mail binnenhalen en een tekst bijwerken?
  - d Zijn er goed gedefinieerde taken die principieel door geen enkele computer perfect kunnen worden uitgevoerd?

## T E R U G K O P P E L I N G

**Uitwerking van de opgaven**

- 1.1 In het adresveld staat nu iets als:  
`https://www.google.nl/search?hl=nl&tbm=isch&q=flamenco`  
Achter de schuine streep staat `tbm=isch` om aan te geven dat we naar afbeeldingen zoeken (Image SearCH). De taal is Nederlands (`hl = nl`) is en we zoeken op het onderwerp flamenco (`q=flamenco`). Als u dit zelf probeert, ziet het adresveld er mogelijk heel anders uit – Google verandert zo af en toe hoe het zoekopdrachten verwerkt.
- 1.2 Webpagina's verwijzen vaak naar elkaar. Als de webcrawler domweg elke link zou volgen, dan loopt deze het risico keer op keer dezelfde pagina op te vragen. Omdat webcrawlers snel zijn, kan dat zeer onaangename gevolgen hebben: als de server aan al die verzoeken tegemoet zou komen, dan raakt deze overbelast en kan niemand meer bij de betreffende pagina. Een slecht ontworpen webcrawler kan dus (onbedoeld) schade aanrichten (al zijn webserver tegen dit soort aanvallen meestal wel bestand; aan te veel verzoeken vanaf dezelfde computer wordt niet steeds voldaan).
- 1.3 In Dijkstra zult u een (uiterst strenge) wiskundige informaticus herkennen. De op interactie gerichte informaticus van Winograd past helemaal niet in deze opsomming. Dit zou kunnen betekenen dat de invloed van deze opvatting uiteindelijk beperkt is, maar het zou ook kunnen betekenen dat Winograd gelijk heeft als hij stelt dat de informatica zelf zich niet aan de nieuwe eisen zal kunnen aanpassen. De toekomst zal dat leren.

**Uitwerking van de zelftoets**

- 1 De zes typen informatici die we hebben besproken zijn de bestuurlijke, de wiskundige, de technische, de managende, de beherende en de gebruikende informaticus. In plaats van de wiskundige informaticus spreken we liever van de kerninformaticus. Voor een korte omschrijving, zie paragraaf 2.1.
- 2
  - a Context van informatica; informatica en samenleving.
  - b Informatiesystemen; modelleren van informatie en processen.
  - c Computersystemen; systeemprogrammatuur.
  - d Softwaresystemen; theoretische informatica.