

Informatica zonder computers

P. Bergervoet

Freudenthal instituut, RU Utrecht

Bij het vak informatica is men vooral geneigd om aan de geautomatiseerde systemen te denken: computers en software. De meeste invullingen van informatica besteden veel aandacht aan het leren omgaan met computers, met softwarepakketten of, als men wat dieper op de informatica in wil gaan, aan het programmeren. Dergelijke invullingen staan dicht bij de computer en bij de bediening ervan.

Tegen dergelijke invullingen zijn verschillende bezwaren aan te voeren. Ze bevinden zich op een hellend vlak. In de informatietechnologie gaan de ontwikkelingen snel. Op dit moment is er WP5.1, volgend jaar is er WP6.0 en over vijf jaar (?) zal WordPerfect als het tekstverwerkingspakket zijn verdwenen. Ook door de gerichtheid op de bediening van het programma hoort WP eerder thuis in een cursus computergebruik (voor nu) dan in het vak informatica.

Ook het belang van leren programmeren in een algemene programmeertaal neemt steeds verder af. Met het verschijnen van krachtig gereedschap wordt dit steeds meer het terrein van specialisten.

In een andere benadering kunnen we informatie en informatieverwerking centraal stellen. Dan gaat het vooral om het analyseren en modelleren van de informatieverwerking en pas in tweede instantie om het gereedschap waarmee het model op de computer vorm krijgt. Bij administratieve toepassingen kunnen we bijvoorbeeld denken aan het beschrijven van de informatie in tabellen en het beschrijven van de functies van de administratie in bewerkingen op die tabellen. Wanneer dit eenmaal gedaan is kan er een gegevensbankpakket worden uitgezocht om de automatisering uit te voeren. Een ander voorbeeld is *knowledge engineering*, al is dit nog minder ver ontwikkeld. Kennis van een expert kan worden gemodelleerd in als-dan-regels, die door een *expert-system-shell* verwerkt kunnen worden. Ook bij het maken van een spreadsheet staat het modelleren centraal.

De ontwikkelingen in de informatietechnologie werken hier precies omgekeerd. Omdat het gereedschap steeds krachtiger en toegankelijker wordt, verschuift de aandacht van het bouwen (programmeren) van een systeem naar het modelleren.

Natuurlijk sluiten deze benaderingen elkaar niet uit. Wanneer leerlingen in een spreadsheet-cursus op een gegeven moment zelf een rekenvel gaan maken, moet er informatie gemodelleerd worden. Na het opstellen van een 'model voor informatieverwerking' zullen we ook een softwarepakket willen gebruiken om die informatieverwerking daadwerkelijk geautomatiseerd uit te voeren. Het is daarom meer een accentverschuiving, maar wel een accentverschuiving met belangrijke gevolgen: bij het vak informatica zal de computer een aantal lessen lang niet aangezet worden.

In 'Informatica Middenbouw'(1) is ruimte gemaakt voor het analyseren en modelleren van informatie. Dit gebeurt bij het hoofdonderwerp van de methode: relationele gegevensbanken. In twee hoofdstukken leren leerlingen de mogelijkheden van relationele gegevensbanken kennen, daarna is er een hoofdstuk gewijd aan informatie-analyse en het ontwerp van gegevensbanken.

In dit stuk zal het vooral gaan over informatie-analyse. Dit blijkt beslist geen eenvoudige zaak te zijn en daarom zijn er verschillende 'informatie-analyse' methoden ontwikkeld. In 'Informatica Middenbouw' wordt gebruik gemaakt van NIAM (Nijssens informatie-analyse methode). Deze wordt in het kort beschreven, waarbij min of meer de lijn gevolgd wordt die ook in 'Informatica Middenbouw' zit. Daarna volgt er een voorbeeld van informatie-analyse in de klas: de analyse van de Top-40.

Relationele gegevensbanken

De informatie-analyse, die in dit artikel de hoofdrol speelt, wordt uitgevoerd met de bedoeling een relationele gegevensbank op te zetten. Relationele gegevensbanken zijn sterk in opkomst als het middel om administratieve informatieverwerking te automatiseren. Dit gaat veel verder dan het opslaan en opzoeken van gegevens alleen. Met behulp van vraagtaalen kan alle verwerking van informatie op een eenvoudige manier beschreven worden. Zo ontstaat er een compleet informatiesysteem.

De toepassingen van relationele gegevensbanken zijn zeer breed. Een complete schooladministratie kan bestaan uit een aantal tabellen (de gegevensbank) plus een

aantal korte programmaatjes in een gegevensbank-vraagtaal om informatievragen te beantwoorden. Wanneer daar nog wat gereedschap bij komt om de uitvoer van gegevens te verzorgen (etiketten maken, gegevens invullen in standaardbrieven) is de geautomatiseerde administratie compleet.

Ook bij de toepassing van statistiek spelen gegevensbanken een belangrijke rol. Bij het onderwerp 'Relationele gegevensbanken en statistiek' in Wiskunde A wordt het programma DATASTAT gebruikt. Dit bestaat eigenlijk uit twee delen: een deel voor het beheer en de verwerking van de statistische gegevens en een deel voor de statistische verwerking. Dit eerste deel is ook een relationele gegevensbank.

Bij 'Informatica Middenbouw' wordt het programma Databaas gebruikt. Hoewel Databaas anders werkt als bijvoorbeeld Datastat, zijn de essentiële bewerkingen in beide programma's dezelfde. Als voorbeeld van die bewerkingen staat hieronder het programma waarmee leerlingen kunnen worden aangemaakt die boeken uit de schoolbibliotheek te lang in hun bezit houden.

```
gebruik   bieb
selecteer status = uitgeleend en uitleendatum < vandaag - 30
koppel    leerling; bieb.leerlingnummer = leerling.leerlingnummer
sorteer   klas
tabel     klas roepnaam naam auteur titel
```

Dit programma laat zich als volgt lezen:

Bekijk de tabel met gegevens over boeken en uitleningen en zoek daarin de boeken die al meer dan 30 dagen zijn uitgeleend. Verbind deze gegevens met de leerlinggegevens door leerlingnummers te vergelijken, sorteer de tabel en druk een tabel met de relevante gegevens af.

Programma's als dit worden in 'Informatica Middenbouw' door leerlingen geschreven.

Het zou te ver voeren om in dit stuk dieper op relationele gegevensbanken in te gaan. Wanneer u geïnteresseerd bent, kunt u terecht in 'Informatica Middenbouw'. Ook is er nu een TELEAC-cursus 'Relationele gegevensbanken en SQL', waarin de meest gebruikte gegevensbank-vraagtaal van dit moment, SQL, behandeld wordt. Daarnaast besteedt deze cursus ook aandacht aan informatie-analyse.

Een goed ontwerp van de gegevensbank is zeer belangrijk. Daarbij is informatie-analyse nodig. Fout of onpraktisch modelleren zal ertoe leiden dat de mogelijkheden van het systeem beperkt zullen worden en kan er zelfs toe leiden dat allerlei functies niet gedaan kunnen worden. De invloed van het ontwerp op de mogelijkheden en de informatie-analyse zullen bekeken worden aan de hand van een voorbeeld: de gegevensbank van videobanden.

De videobanden

Het voorbeeld dat we bekijken gaat over het opzetten van een geautomatiseerde administratie van videobanden op een school. Het is oorspronkelijk afkomstig van een school, waar de ontwikkeling ook min of meer verliep als hier. In 'Informatica Middenbouw' wordt dit voorbeeld ook gebruikt, om de noodzaak van een goede analyse voor het ontwerp duidelijk te maken. De probleemstelling is:

Een school in Nederland heeft in de afgelopen jaren langzaam maar zeker een grote verzameling video-materiaal opgebouwd. Het gaat om meer dan duizend televisie-programma's, films en dergelijke, die verspreid staan over honderden banden.

Om een programma te vinden moet je in een dik schrift opzoeken op welke band het staat. Dat schrift bevat pagina's als in figuur 1.

Omdat er zo veel banden zijn, kost het zoeken van programma's veel tijd. Ook raken banden nogal eens zoek, omdat een leraar ze vergeet terug te brengen. Er zijn ook banden met verouderde programma's, maar omdat niemand overzicht heeft, worden deze nooit gewist. Wanneer er wel wat veranderd wordt, geeft dat in het schrift een kliederboel. Voor het lenen van banden is niks geregeld, het wordt meestal op een kladje bij het schrift gezet. Daarom moet de administratie geautomatiseerd worden. Met dit systeem moet het makkelijk worden om gegevens over programma's te vinden en up to date te houden. Er moeten overzichtjes per vak gemaakt kunnen worden en ook moet geregistreerd kunnen worden waar een band zich bevindt.

BAND: 3111		aanwezig	
VAK: Economie		bij:	
TELLERTIJD: PROGRAMMA:			
1:	0000	30	Economie 10 (ontwikkelingslanden)
2:	1730	30	Economie 11 (economische orde)
3:	2975	35	Levensloop van een bankbiljet
4:	4200	55	Docum. termijnmarkten zilver en cacao
5:	5700	30	OU econ. beleid Europa 1992
6:			

BAND: 1160		aanwezig	
VAK: Duits		bij:	
TELLERTIJD: PROGRAMMA:			
1	350	150	Mutter Courage
2			

figuur 1: twee pagina's uit het schrift van de videobanden.

Zonder enige analyse wordt een opzet voor een gegevensbank gemaakt en uitgevoerd met behulp van een 'elektronische kaartenbak'. In dit eerste ontwerp zijn de kaarten ongeveer gelijk aan de pagina's uit het schrift: iedere kaart bevat alle gegevens over één band. In acht regels worden de gegevens over de programma's gezet. Figuur 2 bevat een voorbeeld van zo'n kaart.

```

bandnummer [ 3111]
sectie [economie ]
docent [ - ]
prog1 [0000 30' Economie 10 (ontw. landen) ]
prog2 [1730 30' Economie 11 (econom. orde) ]
prog3 [2975 35' Levensloop v.e. bankbiljet ]
prog4 [4200 55' Termijnmarkten zilver & cacao ]
prog5 [ ]
prog6 [ ]
prog7 [ ]
prog8 [ ]

```

figuur 2: een kaart uit de eerste versie van de videobanden-gegevensbank.

Al snel blijkt dit ontwerp een grote tekortkoming te bevatten. Het opzoeken van een programma lukt nauwelijks, omdat het op acht verschillende plaatsen kan staan. Het kan alleen door proberen of door ingewikkelde voorwaarden (*prog1=.. of prog2=.. of ...*) gevonden worden. Overzichten maken lukt ook niet, omdat het systeem alleen kaarten kan sorteren en niet de losse gegevens uit *prog1* tot en met *prog8*.

Naar aanleiding van de ervaringen wordt een nieuw ontwerp gemaakt. Het grote verschil is dat per kaart nu de gegevens over een enkel programma worden vastgelegd, in plaats van de gegevens over een band. Een voorbeeld staat in figuur 3.

```

bandnummer [ 3111]
sectie [economie ]
titel [Economie 11 (econom orde) ]
duur [ 30]
tellerstand [ 1730]
docent [ - ]

```

figuur 3: een kaart uit de tweede versie van de videobanden-gegevensbank.

Het zal duidelijk zijn dat de problemen bij het eerste ontwerp nu niet meer optreden. Programma's zijn gemakkelijk te vinden door te zoeken op *titel*. Een overzicht van de programma's van een sectie kan gemaakt worden door te zoeken op *sectie*, te sorteren op *titel* en het afdrucken van een tabel. Ook kunnen gegevens als *duur* en *tellerstand* nu netjes een eigen veld krijgen.

Na verloop van tijd komt er echter een nieuwe tekortkoming aan het licht. Wanneer een docent(e) een programma in de les wil laten zien, moet hij of zij in de gegevensbank aangeven dat de band is uitgeleend, door het invullen van de docentcode bij het veld *aanwezig_bij*. Vaak blijken banden echter niet in de kast te staan, terwijl ze volgens de gegevensbank niet zijn uitgeleend. Toch is het

niet zo dat docenten 'vergeten' hun code in te vullen, zoals blijkt uit de gegevens, die in tabelvorm in figuur 4 staan.

bandnummer	titel	docent
3111	Economie 10 (ontw.landen)		-
3111	Economie 11 (econom. orde)		-
3111	Levensloop v.e. bankbiljet		GB
3111	Termijnmarkten zilver en cacao		-
1160	Mutter Courage		-

figuur 4: inconsistente gegevens.

Volgens de rijen 1, 2 en 4 is band 3111 niet uitgeleend, terwijl in rij 3 staat dat de band 3111 bij docent GB is. De gegevensbank is inconsistent geworden. Deze inconsistentie is een direct gevolg van overtolligheid van gegevens (redundantie). Eén enkel feit, de plaats van een band, kan meer dan eens worden vastgelegd; om precies te zijn evenveel keren als er programma's op de band staan. Je kunt immers niet een programma lenen, maar alleen de complete band.

Om ook dit probleem op te lossen hebben we wat theorie nodig, waarbij een oude bekende om de hoek komt kijken: de verzamelingenleer.

Wanneer we de zaak vanuit een wiskundig standpunt benaderen, komen we terecht bij de verzamelingenleer zoals die in de eerste uitgave van *Moderne Wiskunde* stond: *Een relatie is een verzameling geordende paren van de vorm (x,y) ,*

Dit is precies wat hier ook aan de hand is. De relatie tussen band(nummers) en docent(codes) kan worden beschreven als $\{ (x,y), x \in \text{banden}, y \in \text{docenten} / x \text{ is aanwezig bij } y \}$. Zo'n relatie is een functie in x als iedere waarde van x niet meer dan één keer kan voorkomen. Dit geldt hier voor het bandnummer. Een band kan maar op één plaats zijn, en kan in de lijst dus maar één keer voorkomen. Als we vooruitlopend op de informatie analyse naar figuur 8 kijken, moeten we concluderen dat van de twee bovenste feiten er maar één goed kan zijn.

De theorie rond gegevensbanksystemen stelt nu dat we de structuur van de gegevensbank aan moeten passen aan deze functies. We moeten voorkomen dat de informatie herhaald wordt. Dit kan gedaan worden door tabellen te splitsen, zoals in figuur 5.

We hebben nu dus twee tabellen, één met bandgegevens en één met programmagegevens. We kunnen er voor zorgen dat ieder *bandnummer* in de tabel *banden* maar één keer voorkomt. Daardoor zal de plaats van een band ook maar één keer worden vastgelegd, de redundantie is verdwenen. Merk op dat het bandnummer ook in de andere tabel voorkomt, maar wel met doublures: op één band

programma's

titel	duur	band-nummer	teller-stand
Economie 10 (ontw.landen)	30	3111	0000
Economie 11 (econom. orde)	30	3111	1730
Levensloop v.e. bankbiljet	35	3111	2975
Termijnmarkten zilver en cac	55	3111	4200
Mutter Courage	150	1160	0350

banden

band-nummer	sectie	docent
3111	Economie	GB
1160	Duits	-

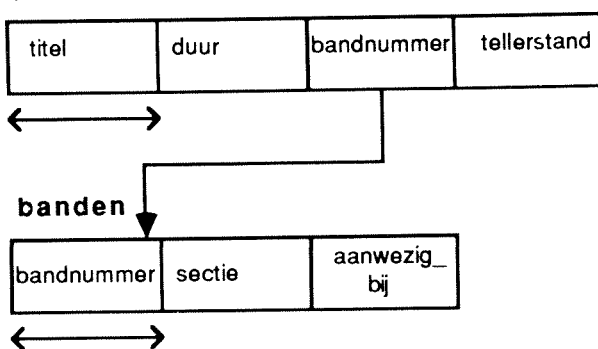
figuur 5: de gegevensbank zonder redundantie.

kunnen immers verschillende titels staan. De bandnummers kunnen nu gebruikt worden om gegevens bij elkaar te zoeken. Het bandnummer in een of andere rij in de tabel *programma's* verwijst naar de rij met hetzelfde bandnummer in *banden*.

Het ontwerp van een gegevensbank met verscheidene tabellen wordt weergegeven in een strokenmodel, zoals in figuur 6. De stroken in dit schema moeten gezien worden als de kopregels van tabellen, met daarin de kolomnamen. De dubbele pijlen onder *titel* en *bandnummer* geven de functies aan: "in deze kolom geen dubbelen". De verwijzing wordt aangegeven door de pijl tussen de tabellen.

Het laatste ontwerp voldoet aan alle eisen die aan het begin van dit voorbeeld gesteld werden. Wel is het zo dat sommige bevragingen wat langer zullen worden, omdat er een koppeling tussen de twee tabellen gemaakt moet

programma's



figuur 6: gegevensbank-schema voor de administratie van videobanden.

worden. Veel voorkomende vragen kunnen echter worden vastgelegd, zodat ze niet steeds opnieuw hoeven worden ingetikt.

Het splitsen van tabellen is op zichzelf niet zo moeilijk, als je maar weet wat de relevante gegevens zijn en wat de relaties tussen deze gegevens zijn. In een praktijksituatie blijkt het erg lastig te zijn om dit vast te stellen. Dit is het terrein van de informatie-analyse.

Informatie-analyse

Het voorbeeld van de videobanden maakt leerlingen duidelijk dat er bij het ontwerpen van een informatiesysteem meer aan de hand is dan het simpelweg schetsen van een kaart voor de kaartenbak. Hiervoor zijn verschillende zogenaamde informatie-analyse methoden ontwikkeld. Een sterk ingekorte versie van één van deze methoden, Nijssens informatie-analyse Methode(2) (NIAM) wordt in het middenbouw-materiaal behandeld.

Het kernpunt van NIAM is dat het volledig gebaseerd is op het gebruik van natuurlijke taal en voorbeelden. Het belangrijkste onderdeel van de analyse bestaat uit een *verwoording van voorbeelden van de relevante informatie*. De informatie wordt eigenlijk opgelezen in zinnen, feiten, alsof het over de telefoon aan iemand verteld moet worden. Er is in principe geen beperking aan het gebruik van natuurlijke taal, behalve één: iedere zin, ieder feit moet een op zichzelf staand stukje informatie geven, onafhankelijk van de context. Met andere woorden, de volgorde van een reeks feiten moet veranderd kunnen worden zonder verlies van informatie.

We zullen ons eens voorstellen dat we als informatie-analist op school binnenkomen met de opdracht een systeem te ontwerpen voor het beheer van de videobanden. De schriftpagina's uit figuur 1 kunnen dan als 'relevante informatie' gezien worden. Een eerste verwoording van die pagina's zou zinnen kunnen bevatten als:

'Levensloop van een bankbiljet' staat op band 3111 van de sectie Economie en begint op 2975 en duurt 35 minuten.

'Mutter Courage' staat op band 1160 van de sectie Duits en begint op 350 en duurt 150 minuten.

....

Band 3111 is aanwezig bij VK.

Band 1160 is aanwezig bij GB.

....

Natuurlijk zijn allerlei verschillende verwoordingen mogelijk. Je zou alle gegevens over band 3111 in één grote zin kunnen stoppen. Verschillen in woordgebruik maken voor het ontwerp niets uit en veel van de andere verschillen zullen bij de volgende stappen in de analyse verdwijnen.

Wanneer op deze manier een flinke portie informatie wordt opgelezen, ontstaan er vanzelf patronen. Als je immers net de gegevens van het ene programma in een zin

hebt gevat, is het alleen maar makkelijk om het volgende programma net zo te doen! Door deze patronen kunnen de variabelen in de zinnen herkend worden. Verder moeten we precies aangeven wat er op deze variabele 'plaatsen' mag worden ingevuld. Op deze manier leiden we uit de feiten de zogenaamde *feittypen* af, de sjablonen voor de informatie. In figuur 7 is dat gedaan.

materie nodig, of er moet iemand bij de hand zijn die vragen kan beantwoorden: "Hebben alle programma's verschillende titels?" In dit voorbeeld nemen we aan dat dit het geval is. Iedere titel komt dus maar één keer voor en bij de splitsing treedt geen verlies van informatie op. Deze (moeilijke) stap levert de relaties die de informatie beschrijven.

Het programma met de TITEL staat op de band met BANDNUMMER van de SECTIE en begint

op TELLERSTAND en heeft een DUUR van minuten.

Economie 10 (ontw.landen)	3111	Economie	0000	30
Economie 11 (econom. orde)	3111	Economie	1730	30
Levensloop v.e. bankbiljet	3111	Economie	2975	35
Termijnmarkten zilver en cacao	3111	Economie	4200	55
Mutter Courage	1160	Duits	0350	150

figuur 7: feittypen met bevolking.

De waarden voor de variabelen, waarmee je de losse feiten weer terugkrijgt, zijn erbij aangegeven.

De volgende stap is het splitsen van de feiten: we proberen de informatie in zo kort mogelijke feiten weer te geven, natuurlijk zonder verlies van informatie. In NIAM worden dit *elementaire feiten* genoemd. Dit is een lastige stap die allerlei vragen oproept, zoals blijkt uit de onderstaande opsplitsing:

Het programma met de TITEL staat op de band met het BANDNUMMER

De band met het BANDNUMMER is van de SECTIE

Het programma met de TITEL begint op TELLERSTAND

Het programma met de TITEL heeft een DUUR van minuten.

De band met het BANDNUMMER is aanwezig bij de docent met DOCENTCODE

Deze opsplitsing is meer dan alleen het in stukken knippen van een zin. Zo moet er goed bekeken worden wat precies bij wat hoort. Zo is de SECTIE aan het BANDNUMMER gekoppeld en niet aan de TITEL. Uit de voorbeeldpagina's kan worden afgeleid dat alle titels op een band toch van één sectie zijn. Ook is het belangrijk om je af te vragen of feiten wel gesplitst kunnen worden. Is de tellerstand van een programma wel los te zien van het nummer van de band waar het programma opstaat? Om dergelijke vragen te kunnen beantwoorden is inzicht in de

Ter vergelijking een paar feiten die niet gesplitst kunnen worden:

De LEERLING E. Beijer heeft voor het CENTRAAL-SCHRIFTELIJK WiskA het CIJFER 6.7 gehaald.

De LEERLING E. Beijer heeft voor het CENTRAAL-SCHRIFTELIJK Engels het CIJFER 7.2 gehaald.

De LEERLING E. Beijer heeft het CENTRAAL-SCHRIFTELIJK WiskA gedaan.

De LEERLING E. Beijer heeft het CENTRAAL-SCHRIFTELIJK Engels gedaan.

De LEERLING E. Beijer heeft het CIJFER 6.7 gehaald.

De LEERLING E. Beijer heeft het CIJFER 7.2 gehaald.

Splitsing leidt tot verlies van informatie: we weten niet meer welk cijfer bij welk vak hoort. Situaties zoals deze zijn typerend voor NIAM. Er wordt steeds naar voorbeelden gekeken en het gebruik van natuurlijke taal zorgt ervoor dat de analyse concreet blijft.

Na het opsplitsen kunnen nog enkele analyse-stappen volgen, zoals het vinden van de afleidbare feittypen (feittypen waarvan de individuele feiten kunnen worden afgeleid uit andere feiten). Een voorbeeld van afleidbare feittypen zal bij het leerlingenwerk ter sprake komen.

De laatste stap in de analyse is het vinden van beperkingen waarin de gegevens in een feittype moeten voldoen. NIAM kent verschillende soorten beperkingen, maar alleen de uniciteitsbeperkingen zijn van essentieel belang voor de analyse. Een uniciteitsbeperking geeft aan dat een gegeven maar één keer voor mag komen bij de feiten van een feittype. Daarmee kunnen we dus vaststellen of een relatie een functie is. Zoals altijd in NIAM bekijken we weer voorbeelden:

De band met BANDNUMMER 3111 is aanwezig bij DOCENT HK.
De band met BANDNUMMER 3111 is aanwezig bij DOCENT VJ.

↑
twee keer hetzelfde bandnummer

De band met BANDNUMMER 3111 is aanwezig bij DOCENT HK.
De band met BANDNUMMER 3133 is aanwezig bij DOCENT HK.

↑
twee keer dezelfde docentcode

figuur 8: uniciteitsbeperkingen?

De bovenste twee feiten kunnen niet allebei goed zijn: Een band kan immers maar op één plaats tegelijk zijn. Met de onderste twee is er niets aan de hand, er staat dat één docent twee banden geleend heeft. De conclusie moet zijn dat het bandnummer maar één keer voor mag komen. De beperkingen worden aangeduid met de dubbele pijl. In figuur 9 is dit gedaan voor alle feittypen.

Het programma met de TITEL heeft een DUUR van minuten.
↔

Het programma met de TITEL staat op de band met BANDNUMMER
↔

Het programma met de TITEL begint op TELLERSTAND
↔

De band met BANDNUMMER is van SECTIE
↔

De band met BANDNUMMER is aanwezig bij DOCENT
↔

figuur 9: de feittypen met uniciteitsbeperkingen.

Tenslotte moet het ontwerp van de gegevensbank nog opgesteld worden. Dit is de eenvoudigste stap: de feittypen met dezelfde uniciteitsbeperkingen worden weer samengevoegd en de bijbehorende tabellen worden de tabellen in de gegevensbank. Dit levert precies het ontwerp in figuur 6.

NIAM in de klas

Informatica is nog steeds een experimenteel vak, alleen met moeite kan er in het programma van een school een beetje ruimte voor gemaakt worden. Omdat informatie-analyse bovendien pas aan het eind van de middenbouw-cursus gedaan wordt, zijn nog weinig scholen aan dit on-

derwerp toegekomen. Op een aantal scholen is informatie-analyse gedaan met een kleine groep leerlingen, vrijwilligers. Deze leerlingen leerden allereerst werken met bestaande gegevensbanken, daarna werd de informatie-analyse ingeleid met het voorbeeld van de videobanden en tenslotte werd de informatie-analyse uitgevoerd op verschillende ontwerp-opdrachten. Deze ontwerp-opdrachten bestaan uit een situatieschets, een formulering van een informatiebehoefte (in de vorm van een aantal vragen die met het systeem beantwoord moeten kunnen worden) en voorbeelden van de benodigde gegevens. Leerlingen kunnen er van uit gaan dat deze voorbeelden significant zijn, dat wil zeggen dat de relevante vragen over uniciteit en splitsbaarheid aan de hand van deze voorbeelden beantwoord kunnen worden. De ontwerp-opdracht van de Top-40 is als bijlage bij dit artikel gevoegd. Deze opdracht werd op de RSG Brokdele in Breukelen onder handen genomen.

De leerlingen begonnen met het oplezen van de informatie, waarbij ze zich concentreerden op de losse regels in de Top-40. De gegevens werden allemaal verbonden aan de titel van de plaat. Er ontstond een lijst feittypen in een soort van vrije notatie:

- 1 De TITEL van de song is
- 2 De song TITEL wordt uitgevoerd door ARTIEST
- 3 De PRODUCER heeft de song TITEL geproduceerd.
- 4 De COMPONIST heeft de song TITEL gecomponeerd.
-

Het eerste feittype is opmerkelijk. Het lijkt aan te geven dat een song moet worden aangeduid met een titel, met andere woorden er wordt een verband gelegd tussen een 'object' in de wereld en de representatie ervan in de gegevensbank.

Hierna werd in sneltreinvaart een rij feittypen opgeschreven die allemaal begonnen met De song TITEL..... De leerlingen namen daarbij niet meer de moeite om ook losse feiten te verwoorden, de analyse leek toch vrij makkelijk te zijn. Er ontstond alleen discussie bij de alarmschijven. Hoe moest dit worden aangeduid? De leerlingen besloten dit te doen met een variabele: wel/niet.

- 9 De song TITEL is een ALARMSCHIJF wel/niet.
- 10 De song TITEL heeft een STIPNOTERING wel/niet.
-
- 15 De song TITEL staat deze week op PLAATS

Op dit moment besloot de docent in te grijpen, door het voorlezen van enkele feiten: "Merkwaardig, hier (de eerste Top-40) staat *De song Tom's Diner staat deze week op plaats 2*, en daar (de andere Top-40) staat *De song Tom's Diner staat deze week op plaats 4*!" De leerlingen reageerden: "Maar dat is een andere week!" Terugverwijzend naar de ontwerp-opdracht werd opgemerkt dat het juist de bedoeling was om Top-40-gegevens over een langere tijd te kunnen gebruiken.

De leerlingen besloten een feittype toe te voegen:

16 Deze week is NUMMER ...

Ook hiervan werden door de docent enkele voorbeelden 'voorgelezen': "Deze week is nummer 37." en "Deze week is nummer 39." Hierdoor werd duidelijk dat de toevoeging van deze feittypen niet tot een verbetering leidde. "Moeten we het weeknummer er dan in zetten?" Zonder (mondeling) antwoord verscheen:

15 De song TITEL staat in WEEKNUMMER op PLAATS

Deze formulering was direct aanleiding om ook de andere feittypen na te lopen, want "misschien moet daar ook een weeknummer bij". Het werd snel duidelijk dat dit bij de stip-noteringen nodig was. Bij de alarmschijven wisten ze het niet direct en de twee Top-40-tabellen werden vergeleken. Platen die in beide voorkwamen, waren in allebei alarmschijf óf geen alarmschijf. Het weeknummer werd daarom niet toegevoegd. Vervolgens werden nog andere feittypen genoteerd:

16 De song TITEL stond in WEEKNUMMER de vorige week op PLAATS

17 WEEKNUMMER ... begint op DATUM ...

18 De song TITEL stond in WEEKNUMMER AANTAL weken in de top-40

Toen de lijst van feittypen af was, werd er gezocht naar afleidbare feittypen. Feittype 18 werd er onmiddellijk uitgelicht. De manier van afleiden werd geformuleerd in termen van de Databaas, het programma waar leerlingen eerder mee gewerkt hadden: "Je kunt het uitzoeken door eerst *zoek* te doen (vind de noteringen van één song) en dan *tel* (bepaal het aantal rijen in het resultaat)". Het feittype werd geschrapt. Het is immers geen informatie die je in de gegevensbank opslaat, maar informatie die je door het systeem laat uitzoeken. De leerlingen zagen echter

niet dat de feittypen 15 en 16 precies dezelfde informatie geven, maar dan één week verschoven.

Het vaststellen van de uniciteitsbeperkingen bracht een nieuw probleem aan het licht. Bij feittype 2 kon geen dubbele pijl onder de titel gezet worden. De Top-40 blijkt (onderling verschillende) platen met dezelfde titel te bevatten. In week 37 staan er platen met de titel 'Ritmo de la noche' op 28 en 29. Ze realiseerden zich onmiddellijk dat dit een kleine ramp was voor de gegevensbank-opzet, omdat je "nooit meer kunt zien wat nou bij wat hoort". Er kwam snel een praktische oplossing: "We geven alle platen een nummer." Het eerste feittype werd veranderd in:

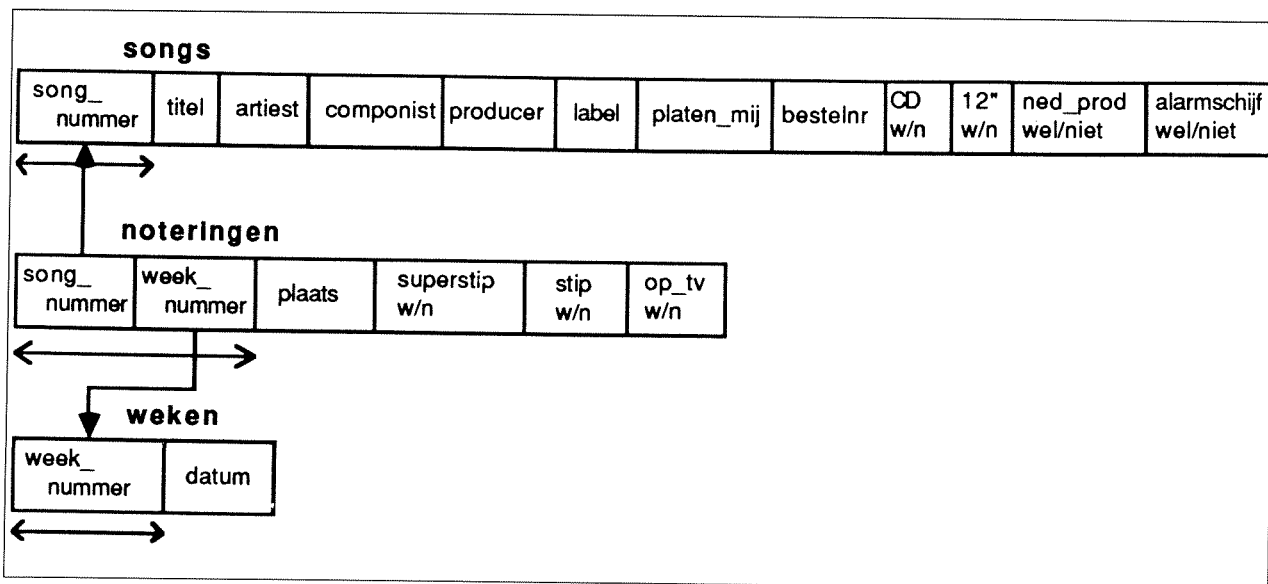
1 De TITEL van de song is en heeft het NUMMER ...

en daarna in:

1 De song TITEL heeft NUMMER

In alle andere feittypen werd de titel in nummer veranderd. Daarna vormden de uniciteitsbeperkingen geen probleem meer. Alle feittypen met alleen het (song)nummer kregen de dubbele pijl onder dit nummer. Daar waar het weeknummer was toegevoegd werd bepaald dat de uniciteitsbeperking voor songnummer én weeknummer gold. Het samenvoegen van de feittypen gaf het gegevensbankontwerp in figuur 10.

Het resultaat heeft nog enkele zwakheden. Door het gebruik van het weeknummer kunnen de noteringen alleen voor één jaar worden vastgelegd. Verder zou de analyse voor de 'songs' nauwkeuriger kunnen. Gegevens over songs en uitvoeringen van songs zouden apart genomen kunnen worden en het voorkomen van meer dan één auteur of producer bij een nummer zou gemodelleerd kunnen worden. Dit is echter niet relevant voor de informa-



figuur 10: strokenmodel van de Top-40-gegevensbank.

tievragen in de opdracht. Voor de Top-40's van één jaar zal dit ontwerp goed werken.

De eerste ervaringen met NIAM lijken aan te geven dat er in het onderwijs met deze methode op een goede manier aan analyse gedaan kan worden. Drie aspecten zijn daarbij van groot belang:

1. NIAM werkt met natuurlijke taal. Daardoor blijft de analyse altijd concreet. De informatie, maar ook de vragen over de structuur van de informatie kunnen altijd in natuurlijke taal geformuleerd worden.
2. NIAM werkt met voorbeelden van informatie. Deze voorbeelden kunnen op een natuurlijke manier gebruikt worden bij reflectie, zoals ook in het voorbeeld van de Top-40 naar voren kwam. Het verwoorden is maar aan zeer weinig regels gebonden, in plaats daarvan komen fouten in de verwoording aan het licht door conflicten met de eigenlijke gegevens.
3. Het is mogelijk gebleken om met een klein deel van NIAM te werken. Dit heeft tot gevolg dat het ontwerpproces niet al te lang wordt en overzichtelijk blijft. Leerlingen worden niet geconfronteerd met een eindeloze serie stappen. De analyse van de Top-40 werd in één les gedaan.

Informatie-analyse in het voortgezet onderwijs?

Tot slot rest de vraag of een onderwerp als informatie-analyse een plaats verdient in het voortgezet onderwijs. Een aantal andere vragen dienen zich daarbij aan. Wat is het nut van informatie-analyse voor leerlingen in het voortgezet onderwijs? Is het onderwerp niet zo specialistisch dat het eerder thuishoort in het beroepsonderwijs? Als we het willen geven, waar moet het dan een plaats vinden?

Naar mijn mening zijn de antwoorden op deze drie vragen respectievelijk "groot", "nee" en "informatica". Ik zal dit toelichten, om te beginnen bij het nut.

In steeds meer vakgebieden en beroepen worden toepassingen van informatietechnologie gebruikt. Vrijwel alle leerlingen zullen in studie of beroep met een of ander pakket te maken krijgen, bijvoorbeeld voor statistische verwerking van gegevens. Ze zullen daarbij ongetwijfeld ook moeilijkheden ondervinden: "Dit kan met pakket X niet gedaan worden" of "We kunnen niet goed genoeg overweg met pakket Y om dit te achterhalen." Vaak is het echter niet zozeer het pakket dat de moeilijkheden veroorzaakt, maar de modellering van de informatie. In de praktijk blijkt bijvoorbeeld dat mensen sterk geneigd zijn om gegevensbestanden te organiseren zoals het eerste voorbeeld van de videobanden. Dit zal de verwerking van

gegevens moeilijk maken. De conclusie is dat het niet zo belangrijk is om te leren werken met pakket X, het is veel belangrijker om te leren hoe informatie gemodelleerd moet worden, zodat informatieverwerking door pakketten van een bepaald type gedaan kan worden⁽³⁾. De specialisatie voor één toepassing kan dan gebeuren wanneer het gebruik aan de orde is.

Naast dit praktisch nut is er naar mijn mening ook een meer algemeen nut. Goed verwoorden en kritisch analyseren van informatie is een belangrijke vaardigheid.

Het antwoord op de tweede vraag ligt in de eerste besluiten. Het is eerder zo dat het leren werken met een bepaald pakket in het beroepsonderwijs thuishoort. Wanneer pakket X bij een bepaald beroep veel gebruikt wordt, is het goed om hier in de beroepsopleiding mee te leren werken. Voor het voortgezet onderwijs is deze beroepspraktijk te ver weg en meer algemene kennis is dan belangrijker.

Bij de vraag naar de plaats speelt een oude discussie een rol. Moet informatica een apart vak worden of moet het worden geïntegreerd in andere vakken. De integratie van onderwerpen als informatie-analyse in andere vakken lijkt mij zeer moeilijk en het gevaar dreigt dat die integratie ingevuld zal worden met computergebruik. Aan de andere kant kan een apart vak informatica zinnig gebruik van informatietechnologie in andere vakken juist stimuleren, omdat in informatica een algemene basis voor dit gebruik gelegd wordt. De computer kan in dit vak informatica een rol spelen als in alle andere: soms kan het apparaat nuttig gebruikt worden en soms kan het ook uit blijven.

Noten

- (1) 'Informatica Middenbouw' is een cursus informatica voor de derde of vierde klassen van havo/vwo. Het materiaal is ontwikkeld in het kader van NIVO en PRINT en, samen met de software, in 1991 uitgegeven bij Educaboek.
- (2) Voor een compleet beeld van NIAM kunt U terecht in de volgende boeken:
Matemataal, Informatie-Analyse en gegevensbank-ontwerp. (Dit boek is mede geschreven voor het voortgezet onderwijs.)
G. Nijssen and T. Halpin. *Conceptual Schema and Relational Database Design.*
- (3) Op dit moment is het gebruik van relationele gegevensbanken in allerlei toepassingen sterk in opkomst. Er wordt gesproken van de 'standaard voor de komende vijftien jaar'. Informatie-analyse zal daarom beslist niet snel verouderen.

Ontwerp-opdracht 1

De Nederlandse TOP 40

Ook in Hilversum raken de gegevensbanken ingeburgerd. Wanneer er een nieuwe plaat wordt aangekondigd, wordt daar tegenwoordig meestal een hoop extra informatie bij gegeven:

De vorige hit van deze artiest was

Deze plaat stond ... weken in de Top-40 en bereikte een hoogste notering van

Je kunt er rustig van uitgaan dat de disk-jockey geen stapels (papieren) top-40's bewaart en deze informatie met de hand opzoekt. Ook de top-40's zitten in een gegevensbank.

Informatievragen

Hieronder staat een aantal opdrachten die met behulp van het informatiesysteem voor de top-40 uitgevoerd moeten kunnen worden.

1. Maak een overzichtje van alle noteringen van één bepaalde plaat.
2. Vind alle platen van een artiest of groep, die ooit in de Top-40 gestaan hebben.
3. Vind van één plaat het aantal weken dat deze in de Top-40 gestaan heeft en de hoogste notering die deze bereikt heeft.
4. Vind van alle platen het aantal weken dat ze in de Top-40 gestaan hebben en de hoogste notering die ze bereikt hebben.
5. Maak een complete afdruk van de top-40 van een bepaalde week.
6. (lastig!) Maak de Top-100 voor een jaar.
(Daarbij mag je aannemen dat platen punten krijgen voor de noteringen, zo kan een eerste plaats 40 punten opleveren, een tweede 39, enz. Je kunt ook aan andere puntenwaarderingen denken.)

Opdracht

Bij dit opdrachtblad zijn (gedeeltelijke) afdrukken van een aantal verschillende Top-40's gevoegd. In plaats daarvan kun je ook recente Top-40-blaadjes gebruiken. Deze kun je als voorbeelden van informatie gebruiken. Omdat er over elke plaat nogal veel gegevens staan, kun je de drie kolommen rechts (12", CD, NL) buiten beschouwing laten

- a Gebruik de NIAM-methode om een ontwerp te maken van een gegevensbank voor het beheren van top-40-gegevens.
- b Beschrijf (in de vorm van reeksen opdrachten voor de Databaas) hoe met deze gegevensbank een antwoord kan worden gevonden op de bovenstaande vragen en opdrachten.

Hint

Zorg ervoor dat je niet maar één Top-40 gebruikt om de feiten af te lezen. Er zijn namelijk veel platen die in beide Top-40's voorkomen. Door te vergelijken kun je zien of je feittypen kloppen!

Uitbreiding

Wanneer je een Top-40 in handen krijgt, is dat eigenlijk al het resultaat van een bevraging van een gegevensbank. De Top-40 wordt namelijk samengesteld door de Stichting Nederlandse Top-40 op basis van 'gegevens van handel en industrie'. Kortom, de verkoopcijfers van een plaat bepalen de plaats op de

week 37 - 15 september 1990

TOP 40

26e jaargang

- stip notering
- superstip
- 📺 alarmschijf
- TV zaterdag bij Veronica

Auteursrechten uitdrukkelijk voorbehouden. Gehele of gedeeltelijke overname in welke vorm dan ook alleen na schriftelijke toestemming. Stichting Ned. Top 40 - Postbus 706 - 1200 AS H. surt. Tel. 035-231647
 DE NEDERLANDSE TOP 40 IS SAMENGESTELD DOOR DE STICHTING NEDERLANDSE TOP 40 UIT GEGEVENS VAN HANDEL EN INDUSTRIË

deze week	vorige week	aantal weken	titel - artiest (producer) comp. - label bestelnr. 7" - platenmij	12	CD	NL	deze week	vorige week	aantal weken	titel - artiest (producer) comp. - label bestelnr. 7" - platenmij	12	CD	NL
1	6	1	I'V BEEN THINKING ABOUT YOU - london beat (m philips) henshall/hems/ea. - anxious 43877 - rca	TV	•	•	21	32	2	TONIGHT - new kids on the block (m starr) m starr/a lancellotti - cbs 6561777 - cbs	📺	TV	•
2	3	5	TOM'S DINER - dna featuring suzanne vega (s addabbo/l kaye) s vega - a&m 3905647 - polydor	📺	•	•	22	31	2	PROMISE ME - beverley craven (p samwell-smith) craven - epic 6559437 - cbs			•
3	2	10	U CAN'T TOUCH THIS - mc hammer (mc hammer) james/miller/mc hammer-capitol 2039257 - emi			•	23	13	11	KILLER - adamski (seal) adamski - mca 2292572737 - wea records			•
4	5	12	MOOI WAS DIE TIJD - corry konings (bert schouten) n munro/c de vries - cnr 1423907 - cnr			•	24	17	7	HOLD ON - en vogue (l mcelroy) en vogue/ea. - atlantic 7567879087 - wea records			•
5	6	6	SUMMER OF '69 - bryan adams (adams/clearmountain) vallance - a&m 3900247 - polydor	TV		•	25	35	2	CORAZON - gerard joling (p de wijn) idem - mercury 8782007 - phonogram			•
6	4	12	IT MUST HAVE BEEN LOVE - roxette (c ofwerman) per gessle-partophone 1363807 - emi bovema			•	26	28	3	WALKING BY MYSELF - gary moore (moore/taylor) rodgers - virgin 113519 - virgin			•
7	9	3	IT'S ON YOU - m. c. sar & the real McCoy (freshline all stars) quickmix/eglitz-zyx rec. 62897 - i. m. c.			•	27	16	6	THIEVES IN THE TEMPLE - prince (prince) idem - paisley park records 5439197571 - wea rec.	📺		•
8	8	4	RITMO DE LA NOCHE - chocolate (maliz/ac 16) lagonda/ea. - teldec 9031711887 - wea rec.			•	28	24	4	RITMO DE LA NOCHE - mystic (mystic) lagonda/castioni/ea. - empire 001 - pwl records			•
9	10	7	MALDON (remix) - zouk machine (g houliery/honore) idem - ariola 112651 - ariola	TV		•	29	33	2	RITMO DE LA NOCHE - lorca (huel/rivalon/ea) lagonda/castioni/ea - carrere 14918 - dureco			•
10	12	3	PRAYING FOR TIME - george michael (g michael) g michael - epic 6561987 - cbs			•	30	34	2	JACK TALKING - d stewart & spiritual cowboys (d a stewart) d a stewart - rca 43907 - rca			•
11	14	4	NEXT TO YOU - aswad (aswad/chris porter) robinson/gaye/ea - island 113558 - ariola	📺		•	31	—	1	LIVIN' IN THE LIGHT - caron wheeler (akiba baby ham/va.) c wheeler/a hall - rca 43939 - rca	📺	TV	•
12	26	2	VERDAMMT, ICH LIEB' DICH - matthias reim (dietrich/reim) idem - mercury 877894 - phonogram	TV		•	32	—	1	CINEMA - ice mc (robxy) idem - zyx records 63097 - imc/erd			•
13	19	3	YEPPA - bzn (beltman) veerman/ea - mercury 8782187 - phonogram			•	33	—	1	SOCA DANCE - charles d lewis (e glem) lewis/gordon smith - polydor 8775087 - polydor	TV		•
14	22	3	NAH NEH NAH - vaya con dios (schouts/klein) idem/balle - ariola 113454 - ariola			•	34	—	1	BANNED IN THE USA - luke feat. 2 live crew (d wright/big tony) b springsteen-bite records 1458037 - cnr			•
15	7	11	CLOSE TO YOU - maxi priest (chung/dunbar/tucker) benson/ea. - 10 rec. 113249 - virgin			•	35	23	7	LA RAZA (la raza mix) - kid frost (tony g) a molina jr/g wilson - virgin 113484 - virgin	📺		•
16	21	3	HE SUZIE - henk wijngaard (fred limpens) d v allena - telstar 4758 - telstar	📺	TV	•	36	—	1	POISON - bell biv devue (dr treeze) elliot straitte - mca rec. 2292572607 - wearecords			•
17	15	8	VISION OF LOVE - mariah carey (m carey/b margulies) idem - cbs 6559327 - cbs			•	37	—	1	DAT BRUINE CAFE - pierre van dam (bart vd post) pruyssenaere/sybrands - indisc 8095 - indisc			•
18	18	5	BLAZE OF GLORY - jon bon jovi (d kortchmar/ea) bon jovi - vertigo 8780807 - phonogram			•	38	30	5	MEGAMIX - technotronic (mastermixers) various - ars/clip records 1457167 - cnr			•
19	11	6	THE WAY YOU DO THE THINGS... - ub 40 (ub 40) robinson/rogers - virgin 113538 - virgin			•	39	25	9	HEAR THE DRUMMER - chad jackson (c jackson) idem - big wave 36 - big wave			•
20	20	4	MOVE TO THE BIGBAND - b liebrand/l scott (b liebrand) b liebrand/l scott - epic 6561767 - cbs			•	40	29	5	LUCILLE - normaal (m hoogenboezem/normal) collins/ea. - cnr 1423957 - cnr			•