

Energie Prestatie Advies

Onderzoek naar het ontwerp en de toepassing van een rapportgenerator
ten behoeve van de geautomatiseerde uitvoering van EPA's

Afstudeerscriptie van Christian Boer (9080285)

Technische Universiteit Delft
Faculteit Informatie Technologie en Systemen
Maart 2003



Technische Universiteit Delft
Faculteit Informatie Technologie en Systemen

Basiseenheid Kennistechnologie
Voorzitter: Prof. Dr. H. Koppelaar
Examinator: Prof. Dr. Ir. E.J.H. Kerckhoffs
Examinator: Ir. A. van den Dool
Examinator: Ir. M. Seepers

Energie Prestatie Advies
Onderzoek naar het ontwerp en de toepassing van een rapportgenerator ten behoeve van
de geautomatiseerde uitvoering van EPA's
door
Christian Boer

Begeleider TU Delft: Prof. Dr. H. Koppelaar
Begeleider BMA Groep: Ir. A. van den Dool
Begeleider BMA Multimedia: R. Schot

AFSTUDEERSCRIPTE

Academiejaar 2002-2003

Voorwoord

Dit rapport is een verslag van mijn afstudeerproject in de studie tot Ingenieur Technische Informatica aan de Technische Universiteit Delft. Dit rapport is in eerste instantie bedoeld als afstudeerscriptie voor mijn studie. Ten tweede is het bedoeld voor de projectmedewerkers van de verscheidene (energie-)projecten die uitgevoerd worden binnen de BMA Groep Ridderkerk.

Op dit punt wil ik een aantal mensen bedanken. Zonder deze mensen had ik nooit tot deze scriptie kunnen komen. Mijn speciale dank gaat uit naar prof. Koppelaar voor zijn uitstekende begeleiding vanuit de universiteit. Verder gaat mijn speciale dank uit naar Richard Schot en Arjen van den Dool van BMA voor het mogelijk maken en begeleiden van dit project.

Andere mensen die ik bij naam wil bedanken zijn Dirk-Jan Los, goede vriend, degene die mij bij de BMA Groep heeft betrokken, Martin van Dord, projectleider Energieloket en manager BMA Advies en Rolf Haaksema, expert EPA en energie binnen BMA Advies.

Als laatste wil ik mijn collega's binnen BMA Multimedia en mijn collega's binnen de BMA Groep Ridderkerk bedanken voor de prettige samenwerking, evenals mijn directe familie en vrienden voor de morele steun tijdens het afstudeerproject.

Christian Boer, maart, 2003

Samenvatting

Energiebesparing krijgt een duidelijke plaats in bestaande wetgeving. Ondernemingen in de Horeca hebben in het algemeen een verbruik van minder dan 100.000 kWh en 50.000 m³ gas, waardoor zij onder de algemene maatregelen van bestuur (AMvB) vallen. Voor de Horeca is dat een speciale AMvB, namelijk een AMvB uit de serie 8.40-besluiten milieubeheer uit december 1998, met de naam "Besluit horeca-, sport en recreatie-inrichtingen". Alle Horecaondernemers met een elektriciteitsverbruik van minimaal 50.000 kWh of 25.000 m³ gas worden geacht te voldoen aan de verplichtingen uit de AMvB.

Een EPA is een integraal energiebesparingadvies voor o.a. ondernemingen. Een EPA rapport beschrijft het pakket aan energiemaatregelen dat een optimale besparing geeft op de energierekening, meer comfort verschaft en het binnenklimaat van het bedrijfspand verbetert. Zo'n EPA rapport voldoet aan alle eisen zoals gesteld in de AMvB.

In het theoretisch gedeelte van deze scriptie ligt de nadruk op het met een computer automatisch genereren van rapporten en de technieken die hiervoor gebruikt kunnen worden. Er wordt daartoe allereerst een overzicht gegeven van bestaande literatuur over tekstgenererende systemen en de daarbij gebruikte AI¹ technieken. De eerste techniek die besproken wordt is de op regels gebaseerde techniek. Met deze techniek kunnen regels vastgelegd worden voor het genereren van zinnen of teksten. Met de Case Based Reasoning (CBR) techniek wordt deze kennis niet opgeslagen in een set van regels, zoals in de eerst besproken op regels gebaseerde techniek, maar in een bibliotheek van uitgewerkte casussen. De CBR-techniek wordt vooral gebruikt in gevallen waarin de te gebruiken kennis niet of nauwelijks vastgelegd kan worden in kennisregels. Vervolgens wordt de Natural Language Generation (NLG) techniek besproken. Deze techniek maakt gebruik grammaticaregels en kennis van taal om teksten te genereren. Vanuit theoretisch oogpunt blijkt er geen verschil in functionaliteit tussen de drie technieken te zijn, met ieder van de technieken kan dezelfde tekst gegenereerd worden. De technieken kunnen echter niet in alle situaties zinvol ingezet worden.

Voor het genereren van de teksten die een rapport vormen wordt op basis van de drie besproken tekstgeneratietechnieken een systeem voorgesteld. De voorgestelde rapportgenerator maakt gebruik van een combinatie van de zinnenplanner van de NLG techniek en een aangepaste op regels gebaseerde techniek. Deze rapportgenerator kan met behulp van rapportmodellen EPA rapporten genereren op basis van informatie verkregen bij de uitvoering van een geautomatiseerde EPA. Deze geautomatiseerde EPA behelst grotendeels dezelfde stappen als de niet-geautomatiseerde uitvoering van een EPA. Een ondernemer kan het systeem gebruiken om vast te stellen of de uitvoering van een EPA voor zijn onderneming verplicht of wenselijk is. De EPA adviseur gebruikt vervolgens het systeem om berekeningen aan de energiehuishouding van de onderneming te maken en een voorlopig EPA rapport te genereren. Er wordt hierbij gebruik gemaakt van het contextgevoelig genereren van tekst met behulp van het gebruikersprofiel. De EPA adviseur zal dit rapport, indien nodig, aanpassen en vervolgens aan de ondernemer presenteren.

Het (software) ontwerp van het systeem voor de geautomatiseerde EPA uitvoering wordt opgesteld met behulp van de modelleringstaal UML. UML staat voor Unified Modeling

¹ Artificial Intelligence (Kunstmatige Intelligentie)

Language en is een grafische taal voor het modelleren en ontwerpen van softwaresystemen, organisaties en andere niet-software gerelateerde systemen. UML bevat een collectie van bewezen modelleringstechnieken voor het modelleren van grote en complexe systemen. UML dekt daarbij alle fasen van het ontwerp, van het analyseren van de specificaties tot het verwezenlijken van de software.

Bij de voorgestelde geautomatiseerde uitvoering van een EPA zijn de ondernemer en de EPA adviseur betrokken. Onderzocht is hoe de rapportgenerator ingezet kan worden in een systeem waarin de ondernemer rechtstreeks zijn EPA rapport krijgt van het systeem. Dit wordt het doe–het–zelf EPA systeem genoemd en is mogelijk te gebruiken in het Energieloket. In deze doe–het–zelf EPA zijn de taken van de EPA adviseur volledig vervangen door een geautomatiseerd systeem. De belangrijkste eis voor een succesvolle invoering van dit systeem is dat de rapportgenerator (nagenoeg) perfect werkt, dus dat er geen verschil is tussen een rapport geschreven door een EPA adviseur en een rapport voortgebracht door een rapportgenerator. Deze eis blijkt echter moeilijk te verifiëren omdat er zelfs tussen de door verschillende EPA adviseurs opgestelde rapporten verschillen zitten.

Conclusies en aanbevelingen.

Inhoud

HOOFDSTUK 1 INLEIDING	- 6 -
1.1 AANLEIDING TOT HET ONDERZOEK.....	- 6 -
1.2 PROBLEEMSTELLING EN DOELSTELLING	- 7 -
1.3 LEESWIJZER	- 8 -
HOOFDSTUK 2 ENERGIE PRESTATIE ADVIES.....	- 9 -
2.1 EPA	- 9 -
2.2 OMGEVING	- 9 -
2.3 PROCESBESCHRIJVING.....	- 10 -
2.4 EPA RAPPORT	- 11 -
HOOFDSTUK 3 THEORIE OVER TEKSTGENERATIE	- 13 -
3.1 TEKST	- 13 -
3.2 OP REGELS GEBASEERDE SYSTEMEN	- 14 -
3.2.1 <i>Eenvoudig teksten genereren met behulp van regels.....</i>	- 14 -
3.2.2 <i>Teksten genereren op basis van numerieke resultaten</i>	- 15 -
3.2.3 <i>Teksten genereren op basis van sjablonen.....</i>	- 16 -
3.3 CASE BASED REASONING	- 18 -
3.4 NATURAL LANGUAGE GENERATION.....	- 20 -
3.4.1 <i>Architecturen.....</i>	- 21 -
3.4.2 <i>Tekstplanner.....</i>	- 23 -
3.4.3 <i>Zinnenplanner (1).....</i>	- 23 -
3.4.4 <i>Zinnenplanner (2).....</i>	- 30 -
3.4.5 <i>Taalkundige Realisator</i>	- 32 -
3.4.6 <i>Een toepassing van NLG: MAGIC</i>	- 32 -
3.5 OVERZICHT	- 35 -
HOOFDSTUK 4 RAPPORTEN GENEREREN	- 37 -
4.1 RAPPORT	- 37 -
4.2 RAPPORTMODEL OF DYNAMISCH RAPPORT?	- 38 -
4.3 INVULLING (1).....	- 40 -
4.4 INVULLING (2).....	- 42 -
4.5 ARCHITECTUUR	- 44 -
HOOFDSTUK 5 GEAUTOMATISEERD EPA SYSTEEM.....	- 46 -
5.1 WAAROM AUTOMATISEREN?	- 46 -
5.2 WAT AUTOMATISEREN?	- 47 -
5.3 PROCESBESCHRIJVING GEAUTOMATISEERDE EPA.....	- 49 -
5.4 DIAGNOSEVRAGEN	- 50 -
5.5 KLANTPROFIEL.....	- 52 -
5.6 VRAGENLIJSTEN.....	- 53 -
5.7 EPA REKENMODEL.....	- 55 -
5.7.1 <i>Kengetallen</i>	- 55 -
5.7.2 <i>Meetwaarden</i>	- 56 -
5.7.3 <i>Besparingsmaatregelen.....</i>	- 57 -
5.7.4 <i>Uitkomsten.....</i>	- 57 -
5.8 RAPPORTGENERATOR	- 59 -

5.9	RAPPORTMODEL VOORBEELD	- 64 -
HOOFDSTUK 6 ONTWERP		- 67 -
6.1	FUNCTIONEEL ONTWERP	- 67 -
6.1.1	<i>Omgeving</i>	- 67 -
6.1.2	<i>Functionaliteit</i>	- 68 -
6.1.3	<i>Stroomschema</i>	- 69 -
6.1.4	<i>Schermschetsen</i>	- 70 -
6.2	GRAFISCH ONTWERP.....	- 71 -
6.3	TECHNISCH ONTWERP	- 72 -
6.3.1	<i>Klassendiagram</i>	- 73 -
6.3.2	<i>Sequentiediagrammen</i>	- 74 -
6.3.3	<i>Databaseontwerp</i>	- 75 -
6.4	REALISATIE	- 76 -
6.5	IMPLEMENTATIE.....	- 76 -
6.6	DISCUSSIE	- 76 -
HOOFDSTUK 7 TOEPASSINGSMOGELIJKHEDEN.....		- 77 -
7.1	UITBREIDING NAAR ANDERE BRANCHES	- 77 -
7.2	ENERGIECENTRUM: DOE-HET-ZELF EPA	- 77 -
HOOFDSTUK 8 CONCLUSIES EN AANBEVELINGEN.....		- 78 -
8.1	CONCLUSIES.....	- 78 -
8.2	AANBEVELINGEN	- 79 -
HOOFDSTUK 9 REFERENTIES.....		- 80 -
BIJLAGE I	BMA	- 82 -

Hoofdstuk 1

Inleiding

Dit hoofdstuk behandelt allereerst de aanleiding tot en de opzet van het uitgevoerde onderzoek. Vervolgens worden de probleemstelling en de deelvragen vastgelegd. Ten slotte wordt de structuur van deze scriptie met behulp van een leeswijzer duidelijk gemaakt.

1.1 Aanleiding tot het onderzoek

Energiebesparing is van groot belang vanwege internationale afspraken over het klimaatbeleid. In december 1997 is in het Japanse Kyoto het protocol onder het Klimaatverdrag afgesloten. Nederland zal na ratificatie gehouden zijn aan een reductiedoelstelling van 6% van de zes broeikasgassen, als bijdrage aan de EU-reductiedoelstelling. Energiebesparing krijgt mede daarom in Nederland een duidelijke plaats in bestaande wetgeving. Ondernemingen in de horeca hebben in het algemeen een verbruik van minder dan 100.000 kWh en 50.000 m³ gas, waardoor zij onder de algemene maatregelen van bestuur (AMvB) vallen. Voor de horeca is dat een speciale AMvB, namelijk een AMvB uit de serie 8.40-besluiten milieubeheer uit december 1998, met de naam "Besluit Horeca-, sport en recreatie-inrichtingen". Alle horecaondernemers met een elektriciteitsverbruik van minimaal 50.000 kWh of 25.000 m³ gas worden geacht te voldoen aan de verplichtingen uit de AMvB.

De inspanning van gemeenten en provincies bepalen in belangrijke mate of de doelen van het klimaat- en energiebesparingsbeleid worden gerealiseerd. In deze nota's wordt aangeduid dat het instrument Wet milieubeheer een belangrijkere rol gaat spelen, zo zullen bedrijven die niet meedoen aan benchmarking of meerjarenafspraken ook energiebesparende maatregelen moeten treffen. Deze maatregelen kunnen o.a. bepaald worden in een Energie Prestatie Advies (EPA). Een EPA advies wordt gegeven door een EPA adviseur.

Vanaf november 2002 wordt door EPA adviseurs gestart met de uitvoering van EPA's en het opstellen van EPA rapporten in de horeca branche. Zo'n rapport wordt gemaakt voor de horecaondernemer die op een heldere en eenvoudige manier wil weten wat op energiegebied bespaard kan of moet worden.

1.2 Probleemstelling en doelstelling

Om de drempel voor het laten uitvoeren van een EPA-onderzoek bij een (horeca-) ondernemer te verlagen wil BMA Advies een systeem aanbieden waarmee de ondernemer zelf kan bepalen of het voor hem nuttig is een EPA onderzoek uit te laten voeren. Dit systeem moet bovendien de EPA adviseur ondersteunen bij het schrijven van EPA rapporten.

Dit leidt tot de volgende probleemstelling:

Welke methoden en technieken kunnen er gebruikt worden om een systeem te maken dat zelfstandig kan bepalen of een ondernemer verplicht is, dan wel verstandig er aan doet om een EPA onderzoek te laten uitvoeren en tevens welke methoden en technieken er vervolgens gebruikt kunnen worden om het systeem een EPA rapport te laten genereren aan de hand van de onderzochte gegevens bij de onderneming?

De volgende doelstelling is gedefinieerd:

Het onderzoeken van ontwerp- en ontwikkeltechnieken voor het met een computersysteem genereren van rapporten en het gebruik van die technieken in het ontwerp van een geautomatiseerd systeem voor het genereren van EPA rapporten.

Het doel van dit onderzoek is het bepalen hoe de tekstuele inhoud van een (EPA-) rapport met behulp van een computergestuurd systeem samengesteld en in Nederlandse zinnen verwoord kan worden.

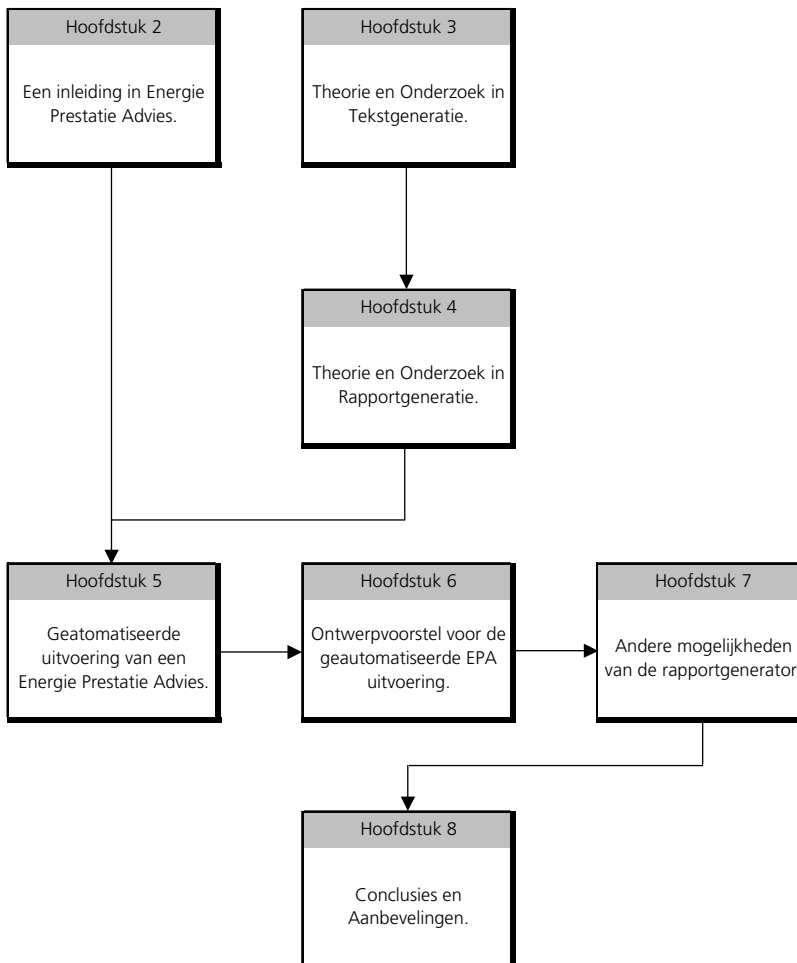
Voor het onderzoek en ontwerp zijn een aantal onderzoeksvragen opgesteld waarin in dit verslag antwoord op gegeven wordt. Deze onderzoeksvragen luiden als volgt:

- Welke methoden en technieken zijn er om teksten te genereren op basis van kennis uit het kennisdomein van een tekstschrijver?
- Hoe kunnen tekstgeneratietechnieken gebruikt worden in een systeem om rapporten te genereren?
- Hoe kan een rapportgenerator toegepast worden voor het genereren van een EPA rapport?
- Welke informatie en berekeningen zijn er nodig om te komen tot een EPA rapport?
- Hoe kan een rapportgenerator zelfstandig ingezet worden om zonder eigen bijdrage van een expert een niet van handgeschreven rapport te onderscheiden rapport te schrijven?

De nadruk van dit onderzoek ligt niet op de zichtbare presentatie van een (EPA-) rapport, dus opmaak en figuren, maar op de vraag met welke technieken de tekstuele inhoud van een EPA rapport vastgelegd en ingevuld kan worden.

1.3 Leeswijzer

Om een computersysteem zelfstandig een EPA rapport te laten genereren, zal allereerst duidelijk moeten zijn wat een EnergiePrestatieAdvies precies inhoudt. Daarom zal in hoofdstuk 2 worden beschreven wat een EPA is en uit welke onderdelen zo'n EPA bestaat. In hoofdstuk 3 wordt het onderzoek naar AI tekstgeneratietechnieken beschreven en wordt bekeken hoe deze al in de praktijk zijn toegepast. Aansluitend wordt in hoofdstuk 4 behandeld hoe deze technieken ingezet kunnen worden bij het genereren van rapporten en wordt dit samengevat tot een rapportgenerator. Hoofdstuk 5 beschrijft vervolgens wat er in de uitvoering van EPA's geautomatiseerd kan worden en hoe de rapportgenerator hierin ingezet kan worden. Het voorgestelde geautomatiseerde systeem dient vervolgens ontworpen te worden. De opzet van dit ontwerp wordt in hoofdstuk 6 beschreven. Het zelfstandig gebruik van de rapportgenerator, dus zonder interventie van de expert, wordt in hoofdstuk 7 onderzocht. Ten slotte volgen in hoofdstuk 8 de belangrijkste conclusies en aanbevelingen wat betreft het uitgevoerde onderzoek.



figuur 1-1 Leeswijzer scriptie

Hoofdstuk 2

Energie Prestatie Advies

In dit hoofdstuk wordt een beschrijving gegeven van de aandachtspunten bij de uitvoering van een EPA en het opstellen van een EPA rapport. Allereerst wordt besproken wat een EPA is en wie er bij de uitvoering van een EPA betrokken zijn. Vervolgens wordt de inhoud van een EPA rapport uitgewerkt.

2.1 EPA

EPA staat voor Energie Prestatie Advies. Een EPA is een integraal energie besparingsadvies voor woningen, wooncomplexen en bedrijven. Een EPA beschrijft het pakket aan energiemaatregelen dat een optimale besparing geeft op de energierekening, meer leefcomfort verschaft en het binnenklimaat van het bedrijfspand verbetert. Het advies richt zich op praktische verbetermaatregelen zoals de aanschaf van een Hr-ketel, spouwmuurisolatie of het plaatsen van een zonneboiler. Een EPA advies wordt gegeven door een EPA adviseur. Hij berekent de huidige energiekwaliteit van een bedrijfspand of bedrijvencomplex aan de hand van een rekenmodel (e-functies) Deze rekenmethoden maken een vergelijking mogelijk van de energiekwaliteit tussen bedrijven van hetzelfde type. De EPA adviseur onderzoekt welke energiebesparende maatregelen passend en effectief zijn, berekent de hoogte van de toekomstige energiebesparing en geeft een indicatie van de verbetering van het binnenklimaat. Ook brengt hij de kosten en de terugverdientijd van de verbetermaatregelen in kaart. Deze scriptie richt zich alleen op ondernemer in de horecabranche. De beschreven methoden en systemen kunnen echter eenvoudig worden uitgebreid naar andere branches.

2.2 Omgeving

De volgende actoren zijn bij de totstandkoming van een EPA betrokken:

- De ondernemer en zijn onderneming;
- De EPA adviseur;
- De brancheorganisatie.

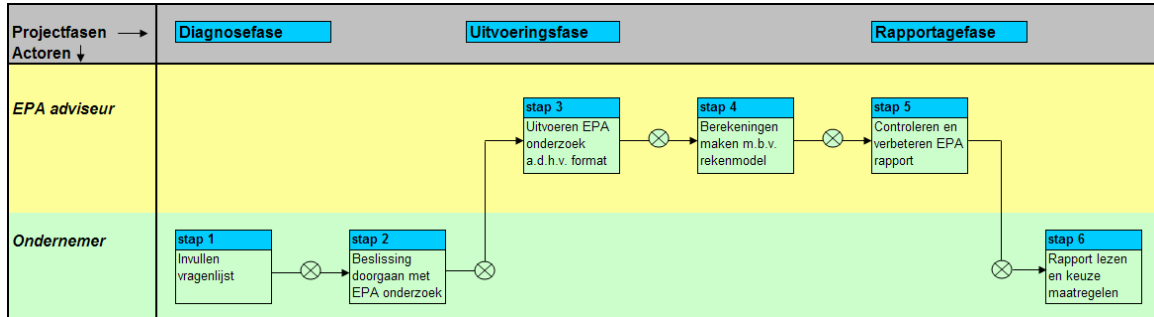
De ondernemer is degene die in zijn onderneming een EPA onderzoek wil laten uitvoeren. Voorbeelden van ondernemingen in de horecabranche zijn o.a. cafés, ijssalons, restaurants en kantines. Zo'n EPA onderzoek zal, indien nodig, uitgevoerd worden door een EPA adviseur. De resultaten van het EPA onderzoek worden in rapportvorm gepresenteerd aan de ondernemer en doorgegeven aan de betreffende brancheorganisatie van de horeca.

2.3 Procesbeschrijving

Een EPA rapport kan in verschillende situaties geschreven worden. De meest voorkomende situatie wordt hieronder beschreven. De chronologische volgorde van acties door de verschillende actoren die bij het totstandkomen van een EPA rapport van belang zijn, luidt als volgt:

- 1) Een ondernemer wil weten of een EPA nuttig is voor zijn onderneming en benadert een EPA adviseur.
- 2) Deze EPA adviseur bepaalt in overleg of de onderneming wettelijk verplicht is een EPA te laten uitvoeren of dat het gezien de situatie waarin de ondernemer zich bevindt verstandig is een EPA te laten uitvoeren.
- 3) Als de ondernemer besluit een EPA onderzoek te laten uitvoeren, zal de EPA adviseur de bedrijfspanden en de energiehuishouding van de ondernemer onderzoeken.
- 4) Met behulp van de meetgegevens uit de vorige stap berekent de EPA adviseur het energieverbruik en rekt hij besparingsmogelijkheden door.
- 5) De EPA adviseur kiest de geschikte rapportvorm.
- 6) Deze besparingsmogelijkheden en bijbehorende terugverdientijden worden in rapportvorm gepresenteerd aan de ondernemer.
- 7) Met behulp van zo'n rapport kan de ondernemer zien hoe zijn energieprestatie is ten opzichte van soortgelijke bedrijven binnen zijn branche en met welke besparingsmogelijkheden hij zijn prestatie kan verbeteren.

Schematisch kan dit op onderstaande wijze worden weergegeven.



figuur 2-1 Schematisch overzicht EPA traject

Het uitvoeren van een EPA traject kan worden onderverdeeld in een drietal fasen. In de diagnosefase wordt contactinformatie over de ondernemer ingewonnen en wordt met behulp van een aantal eenvoudige vragen bepaald of hij in aanmerking komt voor een EPA. De volgende fase betreft de adviesfase waarin een EPA adviseur een uitgebreide doorlichting maakt van de ondernemer. Met de resultaten van deze twee fasen, start de rapportage fase. In deze fase wordt een rapport voor de ondernemer samengesteld.

2.4 EPA Rapport

Zoals beschreven in de vorige paragraaf krijgt de ondernemer aan het einde van het EPA traject een rapport gepresenteerd met daarin de resultaten van het onderzoek. De adviseur beslist over de omvang en inhoud van het rapport aan de hand van de situatie waarin de ondernemer zich bevindt. Hij kan daarbij kiezen uit een viertal verschillende rapportvarianten die variëren in grootte en inhoud. De meeste uitgebreide variant zal 20 á 30 pagina's bevatten, de kleinste ongeveer 3. De globale inhoud van het meeste uitgebreide EPA rapport staat hieronder beschreven. Een EPA rapport bestaat enerzijds uit een aantal standaardteksten dat voor de meeste ondernemers binnen een bepaalde categorie hetzelfde zijn en anderzijds een aantal specifieke teksten dat de EPA adviseur geschreven heeft voor de situatie van de ondernemer. Dit kan bijvoorbeeld een tekst zijn over bepaald apparaat dat de EPA adviseur bij de ondernemer heeft aangetroffen. Globaal zal een EPA rapport de onderstaande inhoud hebben.

INHOUD RAPPORT:

Titelblad

Ieder rapport moet exact hetzelfde voorblad hebben, met uitzondering van de bedrijfsnaam, zodat het voor eenieder herkenbaar is.

Tussenvagina

Alle gegevens van relevante bedrijven en instellingen voor het rapport worden vermeld. Daarnaast staan hierin de plaats, datum, datum van het onderzoek, de contactpersoon van het onderzochte bedrijf en de eigen gegevens van de contactpersoon en adviseur.

Samenvatting

Dit is een korte opsommende samenvatting, waarbij de nadruk ligt op de aanbevolen maatregelen en acties.

Duidelijke inhoudsopgave

De inhoud van het rapport en dus ook de inhoudsopgave is afhankelijk van het uitgevoerde onderzoek, maar zal in de uitgebreide variant de onderstaande hoofdstukken bevatten:

- Voorwoord.
- Samenvatting.
- Hoofdstuk 1: inleiding (huidige problematiek)
- Hoofdstuk 2: een kort theoretisch kader met wet- en regelgeving.
- Hoofdstuk 3: huidige situatie onderneming, dit wordt afgezet tegen de gemiddelde onderneming van dit type om de onderneming te prikkelen beter te worden of te blijven dan de rest.
- Hoofdstuk 4: verslag van de scan met alle mogelijke verbeteropties.
- Hoofdstuk 5: definitieve verbetervoorstellen met op de praktijk gerichte stappen en subsidiemogelijkheden.
- Bijlagen: schematisch weergave met verbeteropties en aandachtspunten voor de onderneming.

Voorwoord

Hierin wordt beschreven voor wie deze rapportage is, wie de opdrachtgever is, een korte aanleiding en een woord van dank. Dit hoofdstuk kan vrijwel in zijn geheel voor ieder bedrijf hetzelfde zijn.

Hoofdstuk 1: inleiding

Een prikkel in de inleiding leidt tot verder lezen. Een minder gemotiveerde ondernemer heeft een meer prikkelende inleiding nodig.

Hoofdstuk 2: theoretisch kader

In dit hoofdstuk wordt ingegaan op de huidige problemen met energiebesparing en de van toepassing zijnde wet- en regelgeving en subsidies. In het rapport wordt duidelijk gewezen op de continu veranderende omstandigheden hierover.

Hoofdstuk 3: huidige situatie ondernemer

In dit hoofdstuk wordt een vergelijking gemaakt tussen de onderneming en het gemiddelde van soortgelijke onderneming. Dit kan een prikkel vormen voor het verbeteren of beter blijven dan collega's en/ of concurrenten. Door bovendien een voorbeeld te gebruiken kan men zien dat de uit het EPA onderzoek naar voren gekomen maatregelen ook daadwerkelijk in de praktijk gebracht kunnen worden.

Hoofdstuk 4: verslag van het EPA onderzoek met verbeteropties

Dit is het belangrijkste deel van het rapport. Hierin staat een korte bespreking van de uitslag van de rekenmethoden en overwegingen om de ondernemer in de betreffende klasse in te delen. Daarna volgt een opsomming van verbeteropties. Er wordt een duidelijk onderscheid tussen wettelijke verplichtingen en aanvullende mogelijkheden gemaakt. In dit hoofdstuk wordt veel bedrijfsspecifieke informatie opgenomen.

Hoofdstuk 5: definitieve verbeteringen met praktijkgerichte tips

Dit hoofdstuk bevat een overzicht waarin de door de ondernemer uit te voeren verbeteropties worden uitgewerkt. Dit zijn de technische, financiële (investering, besparing, terugverdientijd, subsidies) en uitvoeringsaspecten.

Het onderzoek zoals in deze scriptie beschreven richt zich niet op de uiteindelijke vorm waarin het rapport gepresenteerd zal worden, maar bevat een beschrijving van de methoden en technieken om tot de invulling van een EPA rapport te komen met behulp van een geautomatiseerd systeem.

Hoofdstuk 3

Theorie over tekstgeneratie

In dit hoofdstuk wordt een overzicht gegeven van AI (Artificial Intelligence; Kunstmatige Intelligentie) technieken die gebruikt kunnen worden bij het genereren van teksten in natuurlijke taal. AI is de tak van computerwetenschap die probeert menselijke vaardigheden en natuurlijke intelligentie te 'vangen' in computertoepassingen en werkende simulatiemodellen om daarmee de interactie tussen mens en machine te verbeteren. Allereerst worden twee algemene AI technieken besproken die niet specifiek zijn toegespitst op het genereren van teksten, maar er wel voor gebruikt kunnen worden. Ten slotte wordt een pure AI tekstgeneratietechniek besproken en wordt beschreven wat de voor- en nadelen zijn van de verschillende technieken.

3.1 Tekst

Een te genereren tekst kan bestaan uit een enkele zin, of een aantal losse zinnen, eventueel samengevoegd met voegwoorden tot een lopende tekst. Een voorbeeld van zo'n tekst staat hieronder en dient als basis voor de behandeling van de technieken in dit hoofdstuk.

Het koudwatersysteem bestaat uit tapkranen en spoelknoppen. De vaatwasser is aangesloten op een koudwaterleiding. De wc's, die niet zijn voorzien van spoelonderbrekers, hebben een waterreservoir van 20 liter. De tapkranen zijn wel voorzien van doorstroombegrenzers. De gebouwhoogte maakt het nodig om de waterdruk te verhogen met een hydrofoorinstallatie. De bestaande waterdruk in het systeem voldoet tot en met de derde verdieping. Omdat het onderzochte gebouw echter 4 verdiepingen telt, dient de hydrofoorinstallatie ingezet te worden. Verbeterpunten voor de koudwatersysteem zijn het inzetten van spoelonderbrekers voor de wc's. Tevens dient een hydrofoorinstallatie ingezet te worden om de waterdruk in het koudwatersysteem op de juiste hoogte te brengen.

figuur 3-1 Voorbeeld tekstblok

3.2 Op regels gebaseerde systemen

De eerste techniek die hier behandeld zal worden betreft een op regels gebaseerde methode. In het algemeen zijn grote delen van een tekst vaak opgebouwd uit standaardzinnen die door de tekstschrijver op basis van een beslissing in de tekst geplaatst worden.

Een op regels gebaseerd systeem is een expertsysteem² waarmee deze regels vastgelegd kunnen worden en conclusies uit deze regels afgeleid kunnen worden. Zo'n systeem bestaat daarom uit de onderstaande 3 onderdelen.

- Feitenlijst (bevat alle feiten)
- Kennistabel (bevat alle regels)
- Inferentiemotor (bepaalt conclusies aan de hand van feiten en regels)

De inferentiemotor is het hart van het systeem en bepaalt welke regels uitgevoerd moeten worden en in welke volgorde de uitvoering moet gebeuren. De regel met het grootste belang wordt als eerste uitgevoerd. Een regel is belangrijker naarmate er meer feiten nodig zijn om de regel te activeren. Om te voorkomen dat een regel continu geactiveerd blijft op oude feiten, wordt er gebruik gemaakt van refractie. Dit houdt in dat een regel enige tijd nodig heeft voordat hij weer geactiveerd kan worden.

3.2.1 Eenvoudig teksten genereren met behulp van regels

Met behulp van een op regels gebaseerd systeem kunnen heel eenvoudig teksten samengesteld worden. In het onderstaande voorbeeld wordt uitgegaan van een feitenlijst met slechts één feit. Dit kan bijvoorbeeld het resultaat van een bepaalde kwaliteitsberekening zijn. Als dit resultaat de waarde 8 heeft, kan dit met behulp van de volgende kennisregels uit de kennistabel omgezet worden in een zin:

```
ALS resultaat < 2 DAN "U heeft een zeer zwak resultaat behaald."  
ALS 2 =< resultaat < 5 DAN "U heeft een zwak resultaat behaald."  
ALS 5 =< resultaat < 7 DAN "U heeft een redelijk resultaat behaald."  
ALS 7 =< resultaat < 9 DAN "U heeft een goed resultaat behaald."  
ALS resultaat >= 9 DAN "U heeft een uitstekend resultaat behaald."
```

figuur 3-2 Eenvoudige regels

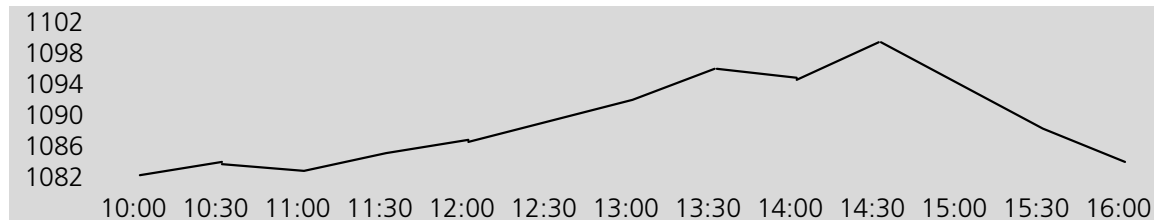
De inferentiemotor zal hieruit "U heeft een zeer goed resultaat behaald." genereren. Deze regels kunnen ook gebruikt worden om tekstblokken te genereren. Zo kan bijvoorbeeld een conclusie bestaande uit meerdere zinnen in het DAN-gedeelte van de kennisregels geplaatst worden.

² Expertsysteem: Programma dat het denkproces van een expert op zijn kennisdomein simuleert.

3.2.2 Teksten genereren op basis van numerieke resultaten

Het vorige voorbeeld behandelde een voorbeeld waarbij een enkele zin gegenereerd werd op basis van vijf regels.

Karen Kukich beschrijft in [5] een ontwerp van een op regels gebaseerd systeem dat een tekstuele rapportage kan geven van koersinformatie. Haar systeem maakt het mogelijk om aan de hand van een tabel met dagelijkse Dow Jones koersgegevens een tekstuele beschrijving te geven van het koersverloop die dag.



figuur 3-3 Dow Jones koersindex

In figuur 3-3 staat het verloop van de koersindex van de 30 grootste aandelen op de New Yorkse effectenbeurs geschetst op 1 december 1983, gebaseerd op een tabel met koersindexhoogtes per half uur. Deze tabel dient als invoer voor het systeem van Kukich dat er de tekst in figuur 3-4 van maakt. Die tekst beschrijft in het Engels de beweging van de aandelenmarkt die dag. Aan de tekst is duidelijk te zien dat het systeem zich hierbij alleen kan baseren op de koershoogtes. Het kan bijvoorbeeld geen uitleg geven waarom vanaf 14:30u zo'n sterke daling optreedt.

after climbing steadily through most of the morning , the stock market was pushed downhill late in the day . stock prices posted a small loss , with the indexes turning in a mixed showing yesterday in brisk trading .

the Dow Jones average of 30 industrials surrendered a 16.28 gain at 4pm and declined slightly, finishing the day at 1083.61 , off 0.18 points .

figuur 3-4 Beschrijving Dow Jones

De globale informatie die nodig is om de daling te kunnen verklaren, bijvoorbeeld een renteverhoging van de centrale bank, is moeilijk in het systeem te vatten. Dit is één van de grootste problemen voor alle tekstgenererende systemen. Het ontbreken van dit soort informatie in de uiteindelijk geproduceerde tekst maakt het echter niet volledig

onbruikbaar. De tekst kan als basis dienen voor de expert die er deze informatie wel aan toe kan voegen. Het systeem is dan meer een intelligente schrijfhulp.

3.2.3 Teksten genereren op basis van sjablonen

De flexibiliteit van de op regels gebaseerde systemen kan nog verder worden uitgebreid door gebruik te maken van sjablonen. Een sjabloon is een standaardzin waar bepaalde woorden ingevoegd moeten worden aan de hand van variabelen. Voorafgaand aan het gebruik van zo'n sjabloon moet met een kennisregel bepaald worden of de zin gegenereerd moet worden. In figuur 3-5 staat een voorbeeld van zo'n sjabloon.

ALS tapkranen EN spoelknoppen DAN
"Het koudwatersysteem bestaat uit tapkranen en spoelknoppen."

figuur 3-5 Sjabloon voorbeeld 1

Dit voorbeeld wil zeggen dat als tapkranen en spoelknoppen gedefinieerd zijn er de volgende zin gegenereerd kan worden door de inferentiemotor:

Het koudwatersysteem bestaat uit tapkranen en spoelknoppen.

waarbij tapkranen 'tapkranen' is en spoelknoppen 'spoelknoppen'.

Zo kan voor iedere zin in het tekstblok van figuur 3-1 een sjabloon gemaakt worden waarmee het tekstblok opgebouwd kan worden. Deze oplossing brengt echter een aantal problemen met zich mee. Het eerste probleem treedt op als alleen tapkranen gedefinieerd is en niet spoelknoppen. In dat geval bestaat het koudwatersysteem dus slechts uit tapkranen. Er is echter geen kennisregel voor een sjabloon die hieraan voldoet zodat deze zin niet gegenereerd wordt. Om de zin wel te genereren moet er een nieuw sjabloon gemaakt worden:

ALS tapkranen EN NIET spoelknoppen DAN
Het koudwatersysteem bestaat uit tapkranen.

figuur 3-6 Sjabloon voorbeeld 2

Met behulp van dit sjabloon wordt de juiste zin wel gemaakt. Dit sjabloon werkt echter niet voor het geval dat spoelknoppen wel bestaat en tapkranen niet. Dus ook hiervoor dient een sjabloon gemaakt te worden.

ALS NIET tapkranen EN NIET spoelknoppen DAN
Het koudwatersysteem bestaat uit spoelknoppen.

figuur 3-7 Sjabloon voorbeeld 3

Energie Prestatie Advies

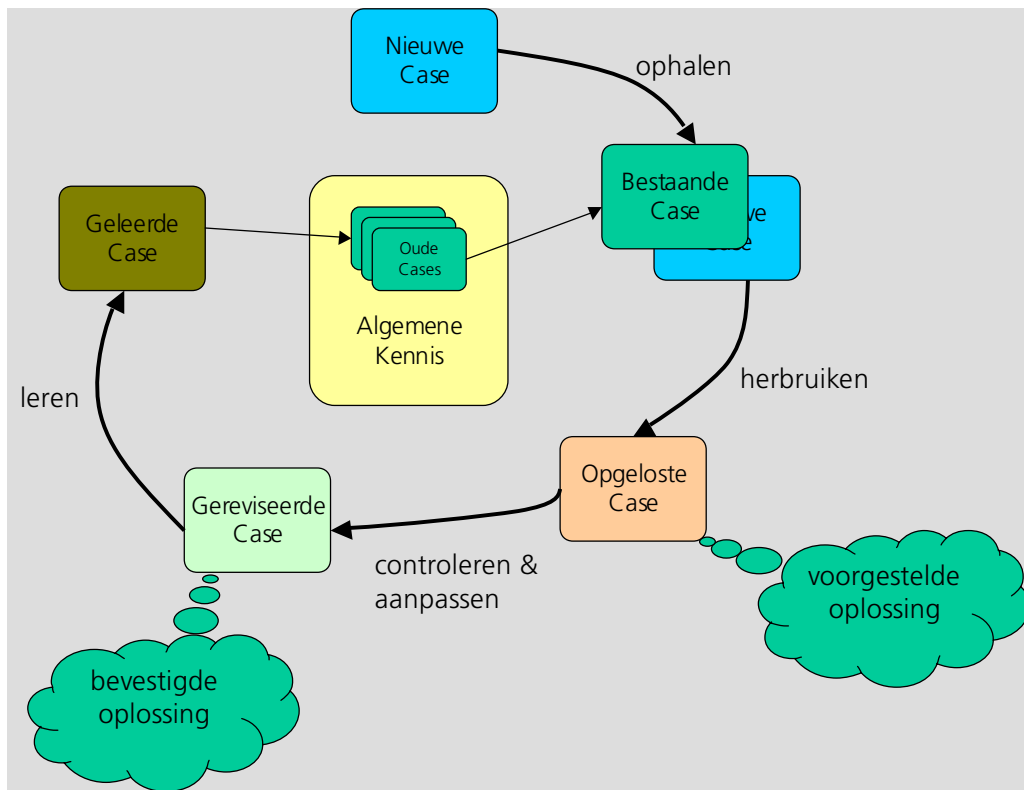
Een nog groter probleem treedt op als het koudwatersysteem van het voorbeeld uit meer dan twee onderdelen bestaat. Dan zou voor iedere mogelijke combinatie een nieuw sjabloon geschreven moeten worden. Voor drie mogelijkheden betekent dit al acht sjablonen.

3.3 Case Based Reasoning

In Case Based Reasoning (CBR; Conclusies trekken op basis van gelijksoortige praktijkgevallen) systemen, wordt kennis opgeslagen in een bibliotheek van uitgewerkte casussen in plaats van een set van klassieke regels zoals in een op regels gebaseerd systeem. Iedere casus bevat een beschrijving van het probleem, plus een oplossing en/of de uitkomst. Het door de expert toegepaste redeneerproces om het probleem op te lossen wordt niet opgenomen, maar zit impliciet in de oplossing.

Case Based Reasoning is het oplossen van nieuwe problemen op basis van de oplossingen van overeenkomstige voorafgaande problemen. Een automonteur die een motor repareert, met behulp van de kennis die hij zich eigen had gemaakt bij de reparatie van een auto van hetzelfde merk, gebruikt Case Based Reasoning. Ditzelfde geldt voor de advocaat die op basis van precedentes zijn strategie voor een nieuwe rechtszaak bedenkt.

Om een probleem op te lossen wordt het probleem vergeleken met bestaande casussen in de casusbibliotheek en gelijke casussen worden opgehaald. De opgehaalde casussen worden gebruikt om een oplossing voor te leggen die opnieuw getest moet worden. Indien nodig kan de oplossing aangepast worden. Ten slotte wordt het huidige probleem en zijn oplossing opgeslagen als deel van een nieuwe casus. Deze procedure is uitgebeeld in figuur 3-8.



figuur 3-8 Case Based Reasoning

CBR systemen worden vooral ingezet in elektronische helpdesks omdat daarbij veelal gelijksoortige problemen opgelost moeten worden. Vaak is de gebruikte kennis niet of nauwelijks expliciet vast te leggen in kennisregels.

Zoals uit de figuur blijkt, bestaat de CBR techniek uit vier stappen. Deze zullen hieronder aan de hand van een voorbeeld toegelicht worden [22].

- Stap 1: Ophalen: In deze stap wordt een bestaande casus opgezocht met de grootst mogelijke gelijkenis met het nieuwe probleem. Zo'n probleem kan bijvoorbeeld zijn dat Fred een kaaspannenkoek wil klaarmaken. Omdat hij nog niet zo vaak heeft gekookt, is de enige ervaring (casus) die hij zich kan herinneren die van een normale pannenkoek die hij ooit heeft gebakken. De procedure die hij daarbij volgde, samen met de toen genomen beslissingen zijn vormen Freds opgehaalde casus.
- Stap 2: Hergebruik: In het voorbeeld van de pannenkoeken moet Fred zijn opgehaalde casus dusdanig wijzigen dat deze ook de toevoeging kaas bevat. Hij gebruikt dus de oude oplossing met een kleine aanpassing eraan.
- Stap 3: Controleren en aanpassen: Nadat de opgehaalde casus is gewijzigd tot de nieuwe oplossing, moet deze oplossing in de praktijk getoetst en eventueel aangepast worden. Stel dat Fred zijn bereidingswijze zo had veranderd dat hij de kaas toevoegt bij het mixen van het beslag. Tijdens het mixen komt hij erachter dat kaas helemaal niet met het beslag wil mengen maar in grote brokken in het beslag blijft drijven, dat was niet de bedoeling. Dit leidt tot de volgende aanpassing: voeg de kaas pas toe als het beslag al in de koekenpan zit.
- Stap 4: Leren: Nadat de voorafgaande oplossing succesvol is aangepast tot de nieuwe oplossing, moet deze als een nieuwe casus in het geheugen opgeslagen worden. Fred zal proberen deze nieuwe oplossing te onthouden en daarmee zijn ervaring in het maken van pannenkoeken te verrijken. Deze ervaring kan hij daarna gebruiken bij het maken van bijvoorbeeld een kersenpannenkoek.

Op het eerste gezicht lijkt CBR sterk op de in de vorige paragraaf geïntroduceerde op regels gebaseerde techniek. Het grootste verschil tussen CBR en de op regels gebaseerde techniek is het tijdstip waarop generalisatie plaatsvindt. Dat wil zeggen in hoeverre nieuwe situaties opgelost kunnen worden. Bij op regels gebaseerde technieken vindt dit plaats tijdens het maken van de regels. In het voorbeeld kan Fred met een op regels gebaseerde techniek geen kaaspannenkoek maken, omdat hij daarvoor de regels nog niet geleerd had. Ditzelfde geldt natuurlijk ook voor de kersenpannenkoek. Voor CBR geldt deze beperking niet. Als Fred eenmaal geleerd heeft, tijdens het koken, hoe hij een kaaspannenkoek moet bakken, kan hij deze kennis ook generaliseren naar een kersenpannenkoek.

Het probleem met op regels gebaseerde technieken is dat het in veel gevallen moeilijk of zelfs onmogelijk is van tevoren alle regels te bepalen. Fred wist toen hij zijn eerste pannenkoek leerde bakken niet dat hij daarna ook nog eens een kaaspannenkoek zou gaan bakken.

3.4 Natural Language Generation

Een meer menselijke manier om teksten te genereren is met behulp van een grammatica. Natural Language Generation (NLG) is een onderdeel van Kunstmatige Intelligentie en Computertaalkunde dat zich bezighoudt met het ontwerp van computersystemen waarmee begrijpelijke teksten in natuurlijke taal gemaakt kunnen worden op basis van een onderliggende niet-taalgebaseerde representatie. Deze computersystemen maken daarbij gebruik van kennis over taal, grammatica en het kennisdomein van de expert om automatisch documenten, rapporten, verklaringen, aanwijzingen en andersoortige teksten te maken.

Er bestaat echter nog een enorm gat tussen door mensen geschreven teksten en door de computer gegenereerde teksten. Er is bijvoorbeeld nog geen enkel NLG systeem in staat om maar enigszins in de buurt te komen van de teksten op de voorpagina van een krant, wat betreft complexiteit en leesbaarheid. Het onderstaande voorbeeld komt uit het Algemeen Dagblad van 9 januari 2003 en is een tekst van 80 woorden. Deze tekst zal door de meeste mensen als niet ingewikkeld ervaren worden, maar is voor een computersysteem nagenoeg onmogelijk om automatisch te genereren.

Alle Nederlandse ambulances en meldkamers worden binnen drie maanden uitgerust met hightech snufjes waardoor beter kan worden bepaald welke ambulances zich waar bevinden.

Bij spoedgevallen kunnen ze dan sneller naar de plek des onheils worden gedirigeerd.

Vice-voorzitter H. Groenendijk van Ambulancezorg Nederland (AZN) verwacht dat ambulances voortaan een minuut eerder ter plekke kunnen komen. De tijdwinst kan worden geboekt doordat elke ambulance wordt uitgerust met een navigatiesysteem. De bestuurder hoeft geen kaart meer te lezen. Patiënteninformatie, indien aanwezig, wordt eveneens meegezonden.

figuur 3-9 Artikel Algemeen Dagblad

Onderzoekers in NLG proberen zulke vloeiende en complexe zinnen te genereren met behulp van computersystemen en hiermee de prestatie van het menselijke brein te evenaren. Om dit te bereiken is het nodig om te identificeren welke informatie gecodeerd moet worden om het systeem aan zulke zinnen invulling te geven en te bepalen welke algoritmen nodig zijn om deze zinnen te construeren.

NLG technologie wordt meestal gebruikt om informatie op eenvoudige wijze aan gebruikers aan te bieden. De technologie kan echter ook gebruikt worden als hulp bij het schrijven van routinedocumenten. Mensen besteden veel tijd aan het schrijven van documenten die vaak grote gelijkenissen vertonen met reeds geschreven documenten. Programma's die daarbij kunnen helpen, kunnen een belangrijke productiviteitsverhoging betekenen.

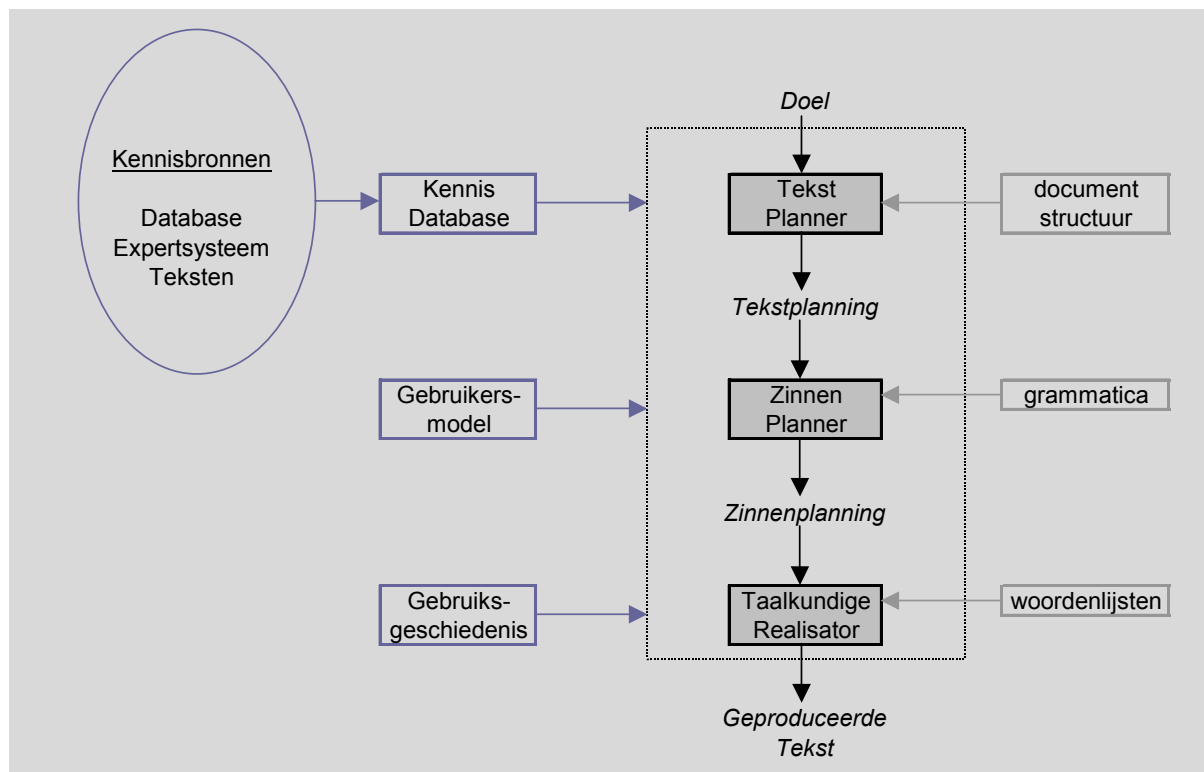
3.4.1 Architecturen

In de jaren 80 was de algemene opvatting onder NLG wetenschappers om het tekstgeneratieproces in tweeën op te splitsen. Dit tweetrapsmodel is opgebouwd uit een strategische en een tactische component die de verscheidene aspecten van het generatieproces op zich nemen. De strategische component bepaalt welke informatie gegenereerd moet worden. De tactische component zet dit om in een leesbare tekst met behulp van zinsaggregatiemethoden. Thompson beschrijft een mogelijke architectuur in [12]. Door de grote complexiteit en het geringe theoretisch begrip van strategische planning was deze architectuur nauwelijks praktisch te realiseren. Bovendien kon domeinafhankelijke kennis nauwelijks gescheiden worden van domeinonafhankelijke kennis, zodat de gebruikte componenten niet of nauwelijks bruikbaar waren voor andere applicaties.

In [2] beschrijven Reiter en Dale daarom een architectuur bestaande uit drie componenten om de genoemde problemen te voorkomen. Zij gaan hierbij uit van de volgende componenten, te weten:

- Tekstplanner;
- Zinnenplanner;
- Taalkundige realisator.

De tekstplanner van Reiter en Dale komt grotendeels overeen met de strategische component van Thompson [11]. De tactische component kan gezien worden als een samenvoeging van de zinnenplanner en taalkundige realisator.



figuur 3-10 Overzicht Architectuur NLG

In deze architectuur wordt het gebruikersmodel los van het doel gezien. Het doel kan bijvoorbeeld zijn om de gebruiker te informeren over een hypotheek. Afhankelijk van zijn inkomen, één van de gegevens in het gebruikersmodel, zal aan de gebruiker een andere hypotheek voorgelegd worden. In welke bewoordingen dit voorstel gedaan zal worden is afhankelijk van de kennis die het systeem van de gebruiker heeft opgedaan in eerdere adviezen. Als daaruit volgde dat de gebruiker meer informatie nodig had over bepaalde termen zal in het nieuwe advies een gedetailleerdere uitleg gegeven worden over mogelijk onduidelijke termen.

In de voorgestelde architectuur wordt van een drietal gebruiker- en domeinonafhankelijk kennisbronnen gebruik gemaakt. De tekstplanner maakt gebruik van een documentstructuur waarin de algemene opbouw van een document beschreven is voor het opbouwen van een domeinafhankelijk tekstplan. Dit tekstplan wordt door de zinnenplanner met behulp van grammaticaregels en het gebruikersmodel omgezet in een zinnenplan. Dit wordt uiteindelijk in de vorm van een leesbare tekst geproduceerd door de taalkundige realisator met behulp van woordenlijsten en kennis van de geschiedenis van de gebruiker.

3.4.2 Tekstplanner

De tekstplanner bepaalt wat de inhoud van de te schrijven tekst wordt en hoe deze tekst opgebouwd en geordend moet worden. De zinnenplanner voegt zinnen samen en kiest de juiste formulering van de begrippen uit de domeinkennis. De taalkundige realisator gebruikt tenslotte de grammaticaregels om syntactisch correcte zinnen te genereren en de tekst op de gebruiker af te stemmen. De invoer van de tekstplanner, het doel, is afhankelijk van de applicatie waarin de NLG techniek gebruikt wordt. De uiteindelijk door de taalkundige realisator geproduceerde tekst kan bijvoorbeeld in de vorm van een HTML of PDF document zijn, of slechts kale tekst zonder opmaak zijn.

3.4.3 Zinnenplanner (1)

De eerste taak van de zinnenplanner is om twee of meer mogelijke berichten in één zin te vatten. Een bericht is een mededeling van het systeem aan de gebruiker en kan bijvoorbeeld een zin zijn. De uitvoer van de tekstplanner wordt gebruikt als invoer voor de zinnenplanner. Dit wordt een tekstplan genoemd en zal een boomstructuur hebben waarbij de bladeren de berichten representeren. De tekstplanner kan bijvoorbeeld met behulp kennisregels of een case based reasoning techniek zijn tekstplan maken. De uitvoer van de zinnenplanner zal een boomstructuur zijn met als bladeren de gecombineerde berichten van het tekstplan. De zinnenplanner zal hierbij moeten beslissen welke berichten samengevoegd moeten worden en op welke manier.

Het invoegen van een zinnenplanner in een tekstgenerator levert volgens Grice in [19] een verbetering van de volgende aspecten op:

- Kracht: Een tekst is krachtiger naarmate hij met minder woorden dezelfde informatie kan uitdragen ("kort maar krachtig"). Dit omvat het verwijderen van overbodige en afleidbare informatie.
- Logica: Een tekst is logischer naarmate hij meer geïntegreerd is. Logica is de eigenschap die van een tekst een logisch geheel maakt in plaats van een warboel van onsamenvangende zinnen.
- Welbespraaktheid: Een tekst is vloeiender naarmate het minder moeite kost de zin te lezen. Factoren die hierbij een rol spelen zijn variatie in zinsstructuur, ondubbelzinnigheid en het vasthouden aan schrijfconventies.

Deze aspecten zijn niet onafhankelijk van elkaar maar hangen nauw samen in een complexe relatie. Bovendien is een aspect als welbespraaktheid nauwelijks objectief meetbaar. Robin wees er in [13] op dat een succesvolle strategie om een tekst krachtiger te maken, het gebruik van complexe zinnen is, hoewel dit paradoxaal lijkt. Dit kan verduidelijkt worden aan de hand van onderstaand voorbeeld:

*De patiënt werd op maandag opgenomen.
De patiënt werd op vrijdag ontslagen.*

Deze zinnen kunnen samengevoegd worden tot:

De patiënt werd op maandag opgenomen en op vrijdag ontslagen.

De samengevoegde versie is krachtiger dan de twee losse zinnen omdat de referenties naar de herhaalde onderwerpen verwijderd zijn. De samengevoegde zin is complexer omdat zij meer informatie uitdraagt in vergelijking met de afzonderlijke zinnen. Halliday en Hasan beschrijven in [20] enkele methoden voor het samenvoegen van zinnen, zoals het gebruik van voornaamwoorden, ellipsen en het verwijderen van redundante gegevens. Deze technieken zullen besproken worden in dit hoofdstuk.

Er zijn meerdere zienswijzen op het samenvoegen van berichten. Reiter en Dale beschrijven in [2] vier verschillende technieken aan de hand van een voorbeeld over treinen. Ze gaan uit van de onderstaande uitvoer van de tekstplanner.

- a. er rijden iedere dag 40 treinen van Amsterdam naar Rotterdam
- b. de volgende trein is de Intercity Plus
- c. de Intercity Plus vertrekt om 10 uur

figuur 3-11 Gedeelte van een tekstplan met drie berichten

1. Voegwoorden

De eenvoudigste manier om twee berichten samen te voegen is het gebruik van het voegwoord *en*. Dit kan gebruikt worden om bijvoorbeeld de eerste twee berichten samen te voegen en heeft het volgende resultaat:

Er rijden iedere dag 40 treinen van Amsterdam naar Rotterdam en de volgende trein is de Intercity Plus. De Intercity Plus vertrekt om 10 uur.

De Nederlandse taal kent naast *en* nog drie groepen met voegwoorden waarvan *of*, *noch* en *maar* de bekendste vertegenwoordigers zijn. Zo kunnen de berichten

*Piet houdt van appels.
Piet houdt niet van peren.*

met het voegwoord *maar* worden samengevoegd tot

Piet houdt van appels, maar Piet houdt niet van peren.

Daarnaast kent de Nederlandse taal een groot aantal varianten dat in één van deze vier groepen geplaatst kunnen worden, zoals *echter*, *zowel ... als*, *niet ... noch*.

In de natuurlijke spreek- en schrijftaal wordt volop gebruik gemaakt van samenvoegen om de verstaanbaarheid te bevorderen. Hiermee wordt bovendien dubbelzinnigheid voorkomen, zoals blijkt uit het onderstaande voorbeeld van Grice in [19]:

*Iemand at appels.
Iemand at bananen.*

Een samengevoegde zin zou de volgende kunnen zijn:

Iemand at appels en bananen.

De normale interpretatie van de eerste twee zinnen zou zijn dat er iemand is die appels eet en een ander die bananen eet. Uit de tweede zin blijkt echter dat het slechts één persoon is die zowel appels als peren eet.

Een andere moeilijkheid betreft de volgorde van voornaamwoorden in een samengestelde zin. Sommige woorden zullen eerder geplaatst moeten worden om de leesbaarheid te bevorderen. De volgende twee zinnen zijn beide grammaticaal correct, maar de tweede is vloeiender dan de eerste:

Harrie at een rode grote appel.

Harrie at een grote rode appel.

Een tekstgeneratiesysteem moet dus proberen deze menselijke woordvolgorde te gebruiken.

Net als deze laatste vervoeging zijn andere voegwoorden die door mensen eenvoudig gebruikt kunnen worden, voor een geautomatiseerd systeem heel wat lastiger te gebruiken. Onderstaand voorbeeld komt uit een schoolles voor kinderen van groep 4 van de basisschool en toont een aantal veelgebruikte voegwoorden.

HERMAN FINKERS

zin 1.= Ik vind Herman Finkers een leuke grappenmaker.

zin 2.= Hij heeft zo'n grappig accent.

voegwoord: omdat

zin 3.= Soms moet ik al om hem lachen.

zin 4.= Hij doet zijn mond open.

voegwoord: zonder dat

zin 5 = Hij liep een keer op het toneel heen en weer.

zin 6.= Je kon zien dat hij wat zocht.

voegwoord: zodat

zin 7.= Hij keek overal.

zin 8.= Hij vond niet wat hij zocht.

voegwoord: maar

zin 9 = Tenslotte was het raak.

zin 10 = Hij deed de klep van de piano open.

voegwoord: toen

zin 11.= Hij had de toetsen gezocht.

zin 12.= Dat was de mop.

Voegwoord: en

zin 13.= Ik heb vier minuten zitten lachen.

zin 14 = Ik had de mop gesnapt

voegwoord: nadat

figuur 3-12 Voorbeeld voegwoorden

De bedoeling is dat de kinderen de zinnen samenvoegen met het juiste voegwoord en in sommige gevallen de volgorde van de woorden in de tweede zin wijzigen. Voor de meeste kinderen is dit niet zo'n moeilijke opgave, omdat zij begrijpen waar een zin over gaat. Voor een geautomatiseerd systeem is dit aan de hand van alleen de zinnen onmogelijk. Hiervoor is kennis nodig over de betekenis van de woorden. Reiter en Dale behandelen slechts de voegwoorden *en* en *maar*.

Wanneer en welke zinnen samengevoegd kunnen worden met deze twee voegwoorden, wordt door Reiter en Dale niet beschreven. Voor de Nederlandse taal kunnen echter een aantal vuistregels gebruikt worden, uitgaande van twee of meerdere zinnen die:

a. beginnen met hetzelfde lidwoord en zelfstandig naamwoord

Het voegwoord *en* kan in het algemeen gebruikt worden om twee of meerdere opeenvolgende zinnen die beginnen met hetzelfde lidwoord en hetzelfde woord samen te voegen. Deze zinnen hebben in het algemeen hetzelfde onderwerp of dezelfde strekking. De onderstaande zinnen kunnen bijvoorbeeld samengevoegd worden:

*De boot arriveert over een uur.
De boot is al onderweg.*

Hieruit ontstaat de grammaticaal correcte zin (zie 2. Ellips voor verdere verbetering):

De boot arriveert over een uur en de boot is al onderweg.

Als in de tweede zin het woord *niet* voorkomt, kan het voegwoord *maar* gebruikt worden in plaats van *en*.

b. beginnen met hetzelfde woord zonder lidwoord

Als twee zinnen beginnen met hetzelfde woord is de kans groot dat deze zinnen hetzelfde onderwerp hebben.

*Piet houdt van appels.
Piet houdt van peren.*

Dit leidt tot:

Piet houdt van appels en Piet houdt van peren.

Deze vuistregel kan ook toegepast worden als de zin niet begint met het onderwerp, maar bijvoorbeeld met een voegwoord:

*Als ik thuis ben, heb ik honger.
Als ik op mijn werk ben, drink ik veel koffie.*

Deze zinnen kunnen samengevoegd worden tot:

Als ik thuis ben, heb ik honger en als ik op mijn werk ben, drink ik veel koffie.

c. hetzelfde onderwerp hebben

Dit is heel wat moeilijker te controleren dan de eerste twee vuistregels. Hiervoor kunnen technieken voor inhoudsanalyse van zinnen gebruikt worden, die echter te ver gaan om hier te behandelen.

2. Ellips³

Als twee samen te voegen berichten een gemeenschappelijke deler hebben, is het soms mogelijk die gemeenschappelijke deler weg te laten uit de samengevoegde zin. Als het tekstplan bijvoorbeeld bestaat uit de zinnen

*Piet ging naar de bank.
Piet stortte € 50.*

dan kunnen deze samengevoegd worden tot

Piet ging naar de bank en stortte € 50.

Het gebruik van ellips kan ook gecombineerd worden met een ander voegwoord uit 1. Zo kan voor het volgende voorbeeld een combinatie van het voegwoord *maar* en ellips gebruikt worden.

*Piet houdt van appels.
Piet houdt niet van peren.*

Samengevoegd leveren deze berichten de volgende zin op:

Piet houdt van appels, maar niet van peren.

De ellips wordt in de Nederlandse taal echter vaak ook uit luiheid gebruikt. Dit levert dan in de meeste gevallen een stijlfout op, zoals in:

Vanochtend vroeg opgestaan, moest helemaal naar Maastricht.

Voor welke zinnen de ellips techniek gebruikt kan worden, wordt opnieuw niet door Reiter en Dale beschreven. De ellips kan heel wat moeilijker toegepast worden dan de eenvoudige voegwoorden. Een mogelijke vuistregel is om twee of meer zinnen die met hetzelfde zelfstandig naamwoord en een werkwoord beginnen samen te voegen, zoals in de eerste twee voorbeelden. Het geautomatiseerde systeem moet dan wel kunnen bepalen of een woord een werkwoord of zelfstandig naamwoord is. De onderstaande zinnen beginnen namelijk ook met twee gelijke woorden, maar kunnen niet met een ellips worden samengevoegd.

*Als ik thuis ben, heb ik honger.
Als ik op mijn werk ben, drink ik veel koffie.*

³ weglating van een woord

Als het systeem geen kennis van de inhoud van de zinnen heeft, kunnen deze foutief worden samengevoegd tot:

Als ik thuis ben, heb ik honger en op mijn werk ben, drink ik veel koffie.

3. Verzameling

Als de samen te voegen berichten identiek zijn met uitzondering van één enkel onderdeel is het vaak mogelijk om deze samen te voegen tot één zin waarin de verschillende onderdelen zijn samengevoegd met een voegwoord. Ter voorbeeld:

*Piet kocht een appel.
Piet kocht een peer.
Piet kocht een banaan.*

Dit kan worden samengevoegd tot:

Piet kocht een appel, een peer en een banaan.

Soms zijn er nog andere domeinafhankelijke verzamelingen mogelijk. Zo kan de opsomming:

maandag, dinsdag, woensdag, donderdag, vrijdag, zaterdag en zondag.

vervangen worden door

alle dagen van de week.

4. Invoeging

Invoeging wordt gebruikt om een bericht als bijzin in een andere zin in te voegen. De relatieve bijzin is waarschijnlijk het bekendste voorbeeld. Zo kunnen de laatste twee berichten van figuur 3-11 op de volgende manieren, afhankelijk van waar de bijzin geplaatst wordt, worden samengevoegd met behulp van invoeging:

De volgende trein is de Intercity Plus, die om 10 uur vertrekt.

of

De volgende trein, die om 10 uur vertrekt, is de Intercity Plus.

Een van de gevaren bij een onzorgvuldige invoeging is het vergeten van de komma voor de bijzin. Dit levert voor de zin uit het voorbeeld op:

De volgende trein die om 10 uur vertrekt, is de Intercity Plus.

Hieruit zou afgeleid kunnen worden dat de Intercity Plus niet de eerstvolgende trein is die vertrekt, maar dat de Intercity Plus de eerste trein is die 10 uur vertrekt en dat er waarschijnlijk nog meer treinen om 10 uur vertrekken.

De moeilijkste taak van de zinnenplanner is het kiezen uit de vele mogelijkheden om berichten samen te voegen. Dit kan enigszins versimpeld worden door de zinnenplanner alleen paren van opeenvolgende berichten samen te laten voegen. Lexicalisatie is het kiezen van woorden om bepaalde domeinbegrippen uit te drukken. Het maken van verwijzingen is meer een beschrijvende taak met als doel het kiezen van woorden om een eenduidige verwijzing naar een begrip in voorgaande zinnen te maken. Dit kan verduidelijkt worden aan de hand van het onderstaande voorbeeld:

De volgende trein is de Intercity Plus. Hij vertrekt om 10 uur.

Met het woord *hij* wordt in de tweede zin verwezen naar de trein in de eerste zin. Het maken van zo'n verwijzing is echter een lastige taak omdat aan het lidwoord *de* niet afgeleid kan worden of er hier sprake van een mannelijk of vrouwelijk zelfstandig naamwoord is. In de volgende zin moet bijvoorbeeld het persoonlijke voornaamwoord *zij* gebruikt worden:

De ontmoeting is geregeld. Zij vindt plaats om 12 uur.

Voor de Nederlandse taal zijn weliswaar vuistregels beschikbaar, maar deze zijn niet sluitend. Veel van deze vuistregels zijn bovendien voor een computer onmogelijk toepasbaar. Het groene boekje [25] biedt een correcte, maar niet-voldoende regel om te bepalen of een woord mannelijk of vrouwelijk is. Deze regel geldt voor alle hieronder genoemde woorden, maar is niet voldoende om een onderverdeling voor alle woorden te maken. Deze regel stelt dat alle woorden met de volgende uitgangen altijd vrouwelijk zijn:

- -heid, -nis, -schap: waarheid, kennis, beterschap;
- -de, -te: liefde, diepte;
- -ij, -erij, -arij, -enij, -ernij: voogdij, rijmelarij, enz.;
- -ing, -st (achter een werkwoordsstam): wandeling, winst;
- -ie, -tie, -logie, -sografie, -agogie: familie, filosofie, demagogie, enz.;
- -iek, -ica: muziek, logica;
- -theek, -teit, -iteit: discotheek, puberteit, subtiliteit;
- -tuur, -suur: natuur, censuur;
- -ade, -ide, -ode, -ude: tirade, periode, enz.;
- -age, -ine, -se: tuigage, discipline, analyse;
- -sis, -xis, -tis: crisis, bronchitis, syntaxis.

Het woord *bank* is een van de woorden die zowel mannelijk als vrouwelijk zijn, net als bloem, kin en stad. Deze woorden zijn van oorsprong vrouwelijk, maar werden in de loop der tijd steeds meer als mannelijk ervaren. In het vorige Groene Boekje, dat van 1954, werden deze woorden nog als vrouwelijk vermeld. In Vlaanderen worden deze woorden meestal nog als vrouwelijk beschouwd en bestaat er dus een voorkeur voor vrouwelijke verwijswoorden:

*De bank en haar medewerkers.
De bloem, zij ligt op de grond.*

In Nederland wordt meestal voor een mannelijk verwijzing gekozen:

*De bank en zijn medewerkers.
De bloem, hij ligt op de grond.*

Volgens de Leidraad van het Groene Boekje zijn de volgende categorieën woorden zowel mannelijk als vrouwelijk:

- De meeste voorwerpsnamen die oorspronkelijk uitsluitend vrouwelijk waren: bank, kast, naald, pijp;
- Algemene aardrijkskundige aanduidingen en namen van hemellichamen: stad, rivier, maan, ster;
- Zelfstandig gebruikte bijvoeglijke naamwoorden en voltooid deelwoorden: zieke, blinde, betrokkene, gewonde;
- Persoonsnamen die voor mannen en vrouwen kunnen worden gebruikt: baby, deugniet, arts, babbelkous.

Verder kent de Nederlandse taal onzijdige zelfstandige naamwoorden. Deze worden altijd mannelijk vervoegd, zoals in onderstaand voorbeeld:

Het huis en zijn bewoners.

3.4.4 Zinnenplanner (2)

Shaw geeft in [11] een andere opsplitsing van zinsaggregatiemethoden, gebaseerd op het type informatie benodigd bij de bewerking. Dit in tegenstelling tot Reiter en Dale die een meer taalkundige opsplitsing aanwenden. Shaw komt tot de onderstaande opsplitsing:

1. Ophelderende samenvoeging

Meerdere berichten worden samengevoegd tot een geringer aantal berichten op basis van domeinspecifieke kennis en bewerkingen.

Dit kan verduidelijkt worden aan de hand van het onderstaande voorbeeld:

*Joop sloeg Kees.
Kees schopte Joop.
Joop schopte Kees.*

Deze drie berichten kunnen samengevoegd worden tot:

Joop vecht met Kees.

2. Verwijzende samenvoeging

De verwijzende samenvoeging wordt door Shaw onderverdeeld in twee varianten. De eerste komt overeen met de invoeging van Reiter en Dale. Een voorbeeld daarvan is:

De volgende trein is de Intercity Plus, die om 10 uur vertrekt.

Het tweede type verwijzende samenvoeging staat bekend als kwantificeren. Bij kwantificeren wordt een verzameling van woorden vervangen door een verwijzing naar hun type met behulp van een kwantor⁴. Deze bewerking staat ook bekend als generalisatie. Een voorbeeld hiervan is:

*Het gaat goed met Joop.
Het gaat goed met Kees.*

Deze twee berichten zouden met behulp van de kwantor *alle* en het type *patiënt* samengevoegd kunnen worden tot:

Het gaat goed met alle patiënten.

Het zal duidelijk zijn dat voor deze bewerking gebruik wordt gemaakt van domeinspecifieke kennis. In dit geval moet bekend zijn dat Joop en Kees patiënten zijn en zelfs dat zij de enige patiënten zijn.

Passonneau gaat hier in [21] verder op in, met onder andere de toepassing van subtypen.

3. Syntactische samenvoeging

De syntactische samenvoeging wordt door Shaw onderverdeeld in de paratactische⁵ en hypotactische⁶ samenvoeging.

Syntactisch gelijke woorden kunnen op paratactische wijze samengevoegd worden. In het onderstaande voorbeeld worden de lijdende voorwerpen appels en peren samengevoegd.

*Joop houdt van appels.
Joop houdt van bananen.*

Joop houdt van appels en bananen.

Bij hypotactische samenvoeging is er sprake van een samenvoeging van twee syntactisch ongelijke onderdelen, zoals in het onderstaande voorbeeld waarin de volgende twee zinnen zijn samengevoegd:

*Joop houdt van Marie.
Marie is de assistente van de tandarts.*

Deze kunnen samengevoegd worden tot:

Joop houdt van Marie, de assistente van de tandarts.

⁴ kwantor: woord of woordgroep waarin informatie gegeven wordt over de hoeveelheid entiteiten waarnaar verwezen wordt.

⁵ parataxis: nevenschikking van woorden met dezelfde, verwante of tegengestelde betekenis.

⁶ hypotaxis: onderschikking.

4. Lexicale samenvoeging

Net als bij hypotactische samenvoeging wordt er bij lexicale samenvoeging gebruik gemaakt van lexicale kennis. In tegenstelling tot hypotactische samenvoeging hoeven de berichten bij lexicale samenvoeging geen voorzetsel of bijvoeglijk naamwoord te zijn.

Voorbeelden van lexicale samenvoeging zijn:

*De dokter is een vrouw → De vrouwelijke dokter.
De dokter is een man → De mannelijke dokter.
Een door de politie gebruikte hond → Een politiehond.*

In tabel 3-1 staan deze vier aggregatiemogelijkheden opgesomd.

Categorie	Bewerkingen	Kennisbronnen	Taalkundig gerealiseerd met
OPHELDEREND	opsommen	<ul style="list-style-type: none">algemene kennisontologie	
VERWIJZEND	verwijzen	<ul style="list-style-type: none">ontologiedomeinkennis	elke, alle, ieder, beide, sommige, een
SYNTACTISCH	parataxis hypotaxis	<ul style="list-style-type: none">syntactische regelswoordenboek	en, die, welke
LEXICAAL	parafraseren	<ul style="list-style-type: none">woordenboek	

tabel 3-1 Overzicht aggregatiemethoden

3.4.5 Taalkundige Realisator

Het laatste onderdeel van de in figuur 3-10 geschetste architectuur is de Taalkundige Realisator. Deze module gebruikt de grammaticaregels om syntactisch correcte zinnen te genereren en de tekst op de gebruiker af te stemmen. Dit afstemmen kan bijvoorbeeld door een toelichting te geven op moeilijke woorden indien uit de gebruiksgeschiedenis is gebleken dat de gebruiker hier moeite mee had.

3.4.6 Een toepassing van NLG: MAGIC

MAGIC (Multimedia Abstract Generation for Intensive Care) is een "intelligent" multimedia presentatiesysteem voor het medische domein [18]. Als een patiënt een hartoperatie heeft ondergaan, moeten de artsen in de operatiekamer (OK) het verplegend personeel in de Intensieve Zorg (IZ) op de hoogte brengen van de resultaten van de operatie, zodat zij de nodige voorbereidingen kunnen treffen als de patiënt naar de IZ gebracht wordt. MAGIC vervangt de rol van de OK-artsen in deze situatie door dezelfde informatie aan te bieden met behulp van automatisch gegenereerde en gecoördineerde spraak en afbeeldingen.

Het leveren van zo'n intern verslag voorafgaand aan de opname in de IZ kan bijdragen aan betere zorgverlening door de verplegers in de IZ. Zij hebben nu meer voorbereidingstijd om de IZ gereed te maken en worden minder vaak geconfronteerd met onverwachte opname van patiënten.

MAGIC maakt gebruik van de uitgebreide online data via het Columbia-Presbyterian Medical Center (CPMC) als bron bij het schrijven van de rapporten. De toestand van de patiënt wordt tijdens de operatie gemeten via het LifeLog database systeem. Hierbij wordt gebruik gemaakt van de meetgegevens van de aanwezige medische apparatuur zoals ventilatoren, hartslagmeters en bloeddrukmeters. De artsen of assistenten voeren tijdens de operatie nog verdere gegevens in. Dit zijn vooral niet-meetbare gegevens zoals beschrijvingen van geluiden van het hart en ademhalingsgeluiden.

Uit deze grote hoeveelheid gegevens haalt het datafilter de voor de IZ relevante gegevens. Als voor ieder feit in het medisch domein een aparte zin gegenereerd wordt, kan de volgende tekst gegenereerd worden:

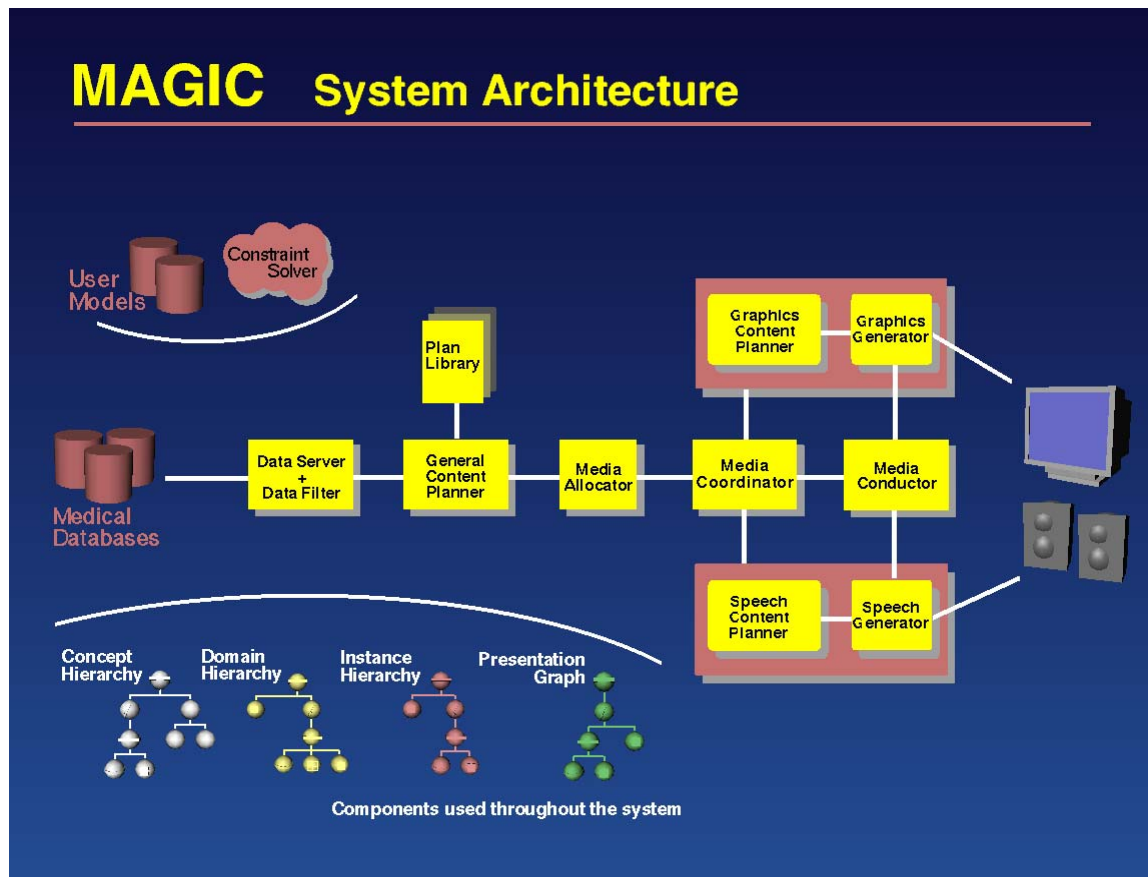
*De patiënt is Jansen.
De patiënt is 80 jaar oud.
De patiënt is een man.
De patiënt heeft ademhalingsmoeilijkheden.
De patiënt heeft diabetes.
De patiënts dokter is Essink.
De patiënt ondergaat een holle ader bypass operatie.*

Het meermaals voorkomen van het woord *patiënt* is duidelijk redundant en zou verwijderd moeten worden. Aggregatieoperatoren kunnen de feiten op duidelijk leesbaardere manier weergeven:

Jansen is een 80 jaar oude, aan ademhalingsmoeilijkheden en diabetes lijdende, mannelijke patiënt van dokter Essink die een holle ader bypass operatie ondergaat.

De zeven zinnen van het eerste voorbeeld zijn vervangen door één complexe zin met dezelfde inhoud. Dit betekent een reductie van 36% in termen van woorden.

Schematisch kan het MAGIC systeem op de onderstaande manier worden weergegeven [6]:



figuur 3-13 MAGIC Architectuur

De MAGIC Architectuur komt grotendeels overeen met de in figuur 3-10 geschetste algemene structuur van een NLG systeem, met een tekstplanner, een zinnenplanner en een taalkundige realisator. MAGIC maakt echter ook nog gebruik van een grafische en spraakmodule om de geproduceerde tekst voor te lezen aan de gebruiker en inzichtelijk te maken met behulp van grafieken en diagrammen.

3.5 Overzicht

In dit hoofdstuk zijn een drietal AI technieken behandeld waarmee teksten gegeneerd kunnen worden. De technieken kunnen echter niet in alle situaties zinvol ingezet worden. Hieronder staat een overzicht van de besproken technieken met hun toepassingsmogelijkheden en voor- en nadelen.

Eigenschap	Techniek		
	<i>Op Regels gebaseerd</i>	CBR	NLG
CONTEXTGEVOELIG GENEREREN VAN TEKSTEN (OP BASIS VAN EEN GEBRUIKERSMODEL)	moeilijk	mogelijk	mogelijk
CONTEXTGEVOELIG AANPASSEN VAN TEKSTEN	moeilijk	moeilijk	mogelijk
GENEREREN VAN NIET VERANDERENDE TEKSTEN	eenvoudig	eenvoudig	eenvoudig
TEKSTEN OP BASIS VAN (INGEVOERDE) GEGEVENS GENEREREN	eenvoudig	eenvoudig	mogelijk
TEKSTEN OP BASIS VAN ONBESCHIKBARE GEGEVENS GENEREREN	onmogelijk	onmogelijk	onmogelijk
VEREISTE KENNIS UIT HET KENNISDOMEIN BIJ ONTWIKKELING VAN TECHNIEK	enigszins	weinig	diepgaand
DYNAMISCHE TEKSTEN VAN 'HOGE KWALITEIT' GENEREREN	moeilijk	moeilijk	mogelijk
ONDERHOUDEN VAN BENODIGDE KENNIS BIJ HET GENEREREN VAN TEKSTEN	mogelijk	impliciet	moeilijk
COMPLEXITEIT VAN HET ONTWIKKELDE SYSTEEM	eenvoudig tot complex	mogelijk tot complex	complex indien geen gebruik gemaakt van bestaande NLG componenten
VERSPREIDING VAN TECHNIEK	relatief veelgebruikt	weinig gebruikt	nauwelijks gebruikt
TOEPASSING VAN DE TECHNIEK	commercieel academisch	voornamelijk academisch	academisch

tabel 3-2 vergelijking technieken

Uit het overzicht blijkt dat NLG en in mindere mate CBR, hoewel beide al meer dan 20 jaar in omloop [22], nog nauwelijks commercieel worden toegepast. Case Based Reasoning wordt vooral in grote projecten ingezet, zoals voor de helpdesk van Compaq [24]. Dit wil echter niet zeggen dat CBR in kleinere projecten niet te gebruiken is. Delen van de CBR techniek, zoals het ophalen van cases, kunnen wel goed ingezet worden binnen kleinere projecten.

Deze drie technieken zijn niet in alle gevallen voldoende om de juiste teksten te genereren. Sommige informatie kan beter op een grafische manier gepresenteerd worden, zoals in het MAGIC voorbeeld. In andere gevallen is het goedkoper om mensen in te huren om documenten te (her-)schrijven, of is het resultaat van door mensen geschreven teksten significant beter. Er zijn geen vaste regels om te bepalen of en welke van deze drie technieken gebruikt moet worden om informatie aan de gebruiker te presenteren. Dit is zowel afhankelijk van de informatie die gepresenteerd moeten worden als de ervaring van de gebruiker met de informatie.

Vanuit een theoretisch oogpunt is er geen verschil in functionaliteit tussen de drie technieken. Met de genoemde technieken kunnen dezelfde zinnen gemaakt worden. Het gebruik van sjablonen in een op regels gebaseerd systeem kan zelfs gezien worden als een eenvoudige vorm van NLG. Pure NLG technieken hebben echter wel als voordeel dat ze teksten van hogere kwaliteit kunnen produceren, vooral in gevallen waar veel variatie in de tekst nodig is. Pure NLG systemen kunnen in gevallen waarbij de tekst en de structuur van de tekst regelmatig verandert betere onderhoudsmogelijkheden bieden. Met NLG gemaakte teksten en teksten die met behulp van sjablonen zijn gemaakt, hebben als voordeel t.o.v. handmatig geschreven teksten dat ze toegespitst kunnen worden op de persoonlijke behoefte van de gebruiker en rekening houden met zijn of haar:

- Aanwezige kennis;
- Behoeften;
- Capaciteiten;
- Situatie;
- Taal.

Op financieel vlak kunnen de in deze paragraaf beschreven technieken voordelen bieden, omdat teksten altijd gebaseerd zijn op de laatste versie van kennis in de kennisdatabase en bestaande documenten dus niet handmatig aangepast hoeven te worden met gewijzigde kennis.

Hoofdstuk 4

Rapporten genereren

In het vorige hoofdstuk zijn AI softwaretechnieken voor het genereren van tekst besproken. Dit hoofdstuk beschrijft hoe deze technieken gebruikt kunnen worden bij het genereren van rapporten. Daartoe wordt allereerst gesteld aan welke eisen een algemeen rapport moet voldoen en uit welke onderdelen zo'n rapport bestaat. Deze onderdelen worden vervolgens behandeld en de daarbij te gebruiken technieken worden samengevoegd tot een architectuur voor het genereren van rapporten.

4.1 Rapport

Alvorens een beschrijving te geven van rapportgenererende technieken is het nodig vast te stellen wat er precies met een rapport bedoeld wordt. In het algemeen kan gesteld worden dat een eenvoudig rapport bestaat uit de volgende elementen:

- Hoofdstukken.
- Paragrafen.
- Subparagrafen.
- Opsommingen.
- Tekst.
- Figuren.
- Tabellen.

Onderdelen als voorwoord en conclusie worden hier niet genoemd, omdat deze opgebouwd kunnen worden uit een combinatie van de genoemde elementen. Andere aspecten zoals kop- en voetteksten worden hier ook niet besproken omdat dit echte opmaakaspecten zijn en niet met behulp van een tekstuele rapportgenerator geproduceerd hoeven te worden. Ook op de invulling van figuren en tabellen wordt hier niet verder ingegaan. Dit is namelijk ook geen taak van de tekstgenerator.

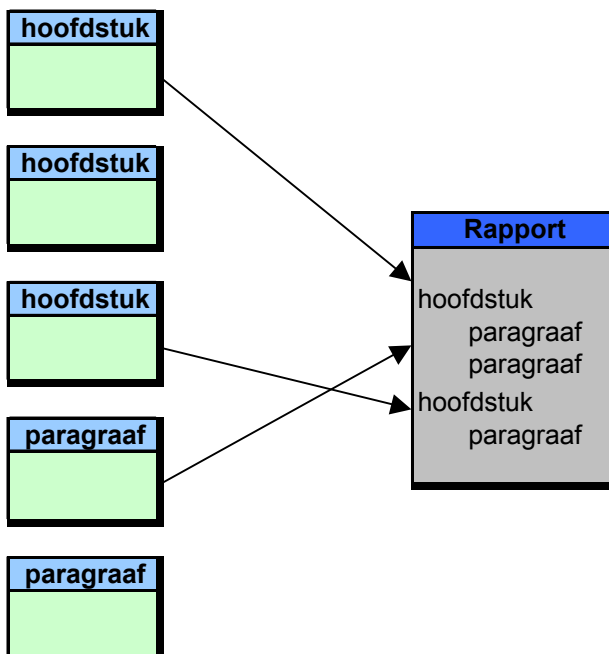
Uitgaande van de bovenstaande beschrijving van een rapport, kunnen de volgende eisen aan een rapport genererend systeem (rapportgenerator) gesteld worden:

- Het systeem moet de inhoud van het rapport op hoofdlijnen kunnen bepalen, dus de indeling in hoofdstukken, paragrafen en standaardteksten kunnen realiseren.
- Het systeem moet het rapport met tekstblokken, figuren en tabellen kunnen invullen aan de hand van gestelde criteria.
- Het systeem moet teksten kunnen samenstellen uit losse zinnen, waarbij iedere zin afhankelijk is van bepaalde criteria en gebruik maakt van bepaalde variabelen.

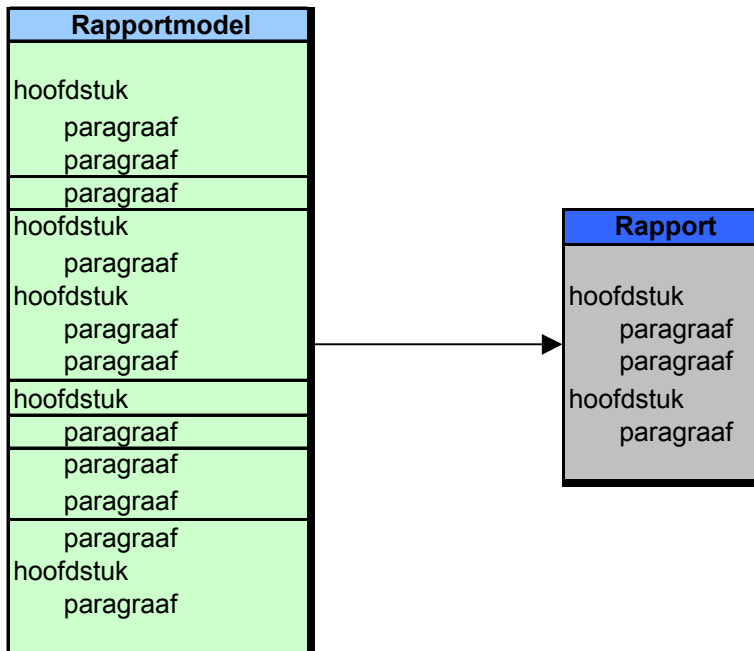
4.2 Rapportmodel of dynamisch rapport?

Zoals gezegd moet het beoogde systeem allereerst de inhoud van het rapport op hoofdlijnen kunnen bepalen. In deze eerste stap wordt een soort inhoudsopgave gemaakt aan de hand waarvan de rest van het te maken rapport ingevuld gaat worden. Hiervoor is een tweetal mogelijkheden te bedenken. De eerste en meest complexe methode gaat uit van een volledige dynamische inhoud van het rapport. Dat wil zeggen dat op basis van bepaalde criteria beslist wordt uit welke hoofdstukken het rapport moet bestaan en welke paragrafen ieder hoofdstuk moet bevatten. Op deze manier kunnen zeer gevarieerde rapporten gegenereerd worden. Het probleem is echter dat van ieder hoofdstuk bekend moet zijn aan de hand van welke criteria het hoofdstuk eventueel in het rapport geplaatst moet worden. Tevens dient van ieder hoofdstuk zijn volgorde bij het plaatsen van de hoofdstukken bekend te zijn, om te voorkomen dat bijvoorbeeld de inleiding achteraan geplaatst wordt.

In de hierboven genoemde methode wordt uitgegaan van een volledig dynamisch rapport op basis van hoofdstukken en paragrafen. Als van tevoren echter de opbouw van het rapport bekend is, kan hiervan handig gebruik gemaakt worden om deze eerste stap bijna volledig over te slaan. Hierbij wordt uitgegaan van een vaste opbouw van een volledig rapport in vooraf gedefinieerde hoofdstukken en paragrafen. Dit kan bijvoorbeeld omdat het systeem altijd dezelfde soort rapporten moet genereren, die qua opbouw gelijk zijn. Ook als er slechts een klein aantal verschillende varianten in de te genereren rapporten zijn, kan deze variant gebruikt worden. In het kort komt het er op neer dat er een rapportmodel opgesteld wordt dat alle mogelijke teksten bevat. Voor ieder deel, bijvoorbeeld een hoofdstuk, een paragraaf, of een tekstblok van dat volledige rapport is aangegeven onder welke voorwaarden het in het uiteindelijke rapport terecht moet komen. In tegenstelling tot de eerst genoemde methode worden er dus teksten verwijderd in plaats van bijgeplaatst. Dit staat geschetst in figuur 4-1 en figuur 4-2.



figuur 4-1 Methode 1 (Dynamisch rapport)



figuur 4-2 Methode 2 (Rapportmodel)

Methode 2 heeft als voordeel dat de volgorde van de hoofdstukken impliciet al vastgelegd is en niet per hoofdstuk bepaald hoeft te worden. Bovendien is het voor veel experts makkelijker om uit te gaan van een volledig samengesteld rapport en daarin aan te geven wanneer welke delen van dit rapport nodig zijn dan een verzameling van losse teksten aan te leggen om daarmee een rapport samen te stellen. Uiteindelijk kunnen met beide methoden dezelfde rapporten gemaakt worden. Beide methoden leveren een rapport dat onderverdeeld is in hoofdstukken en paragrafen met vaste stukken tekst.

4.3 Invulling (1)

Naast de opdeling van het rapport in hoofdstukken en paragrafen, moet het rapport gevuld worden met allerlei teksten, tabellen en figuren. Hiervoor kunnen de in het vorige hoofdstuk beschreven technieken toegepast worden, te weten:

- Op regels gebaseerd.
- Case Based Reasoning.
- Natural Language Generation.

De keuze voor de juiste techniek voor het genereren van de teksten is niet eenvoudig. Bij de keuze van de techniek moet bekend zijn welke informatie de uiteindelijke tekst moet uitdragen, waar de data om de tekst op te bouwen beschikbaar is en hoe deze tekst aan de gebruiker gepresenteerd moet worden. Deze problemen kunnen aan de hand van een voorbeeld verduidelijkt worden. In dit voorbeeld staat een tekst zoals die kan voorkomen in de rapportage aan een ondernemer.

Het koudwatersysteem bestaat uit tapkranen en spoelknoppen. De vaatwasser is aangesloten op een koudwaterleiding. De WC's, die niet zijn voorzien van spoelonderbrekers, hebben een waterreservoir van 20 liter. De tapkranen zijn wel voorzien van doorstroombegrenzers. De gebouwhoogte maakt het nodig om de waterdruk te verhogen met een hydrofoorinstallatie. De bestaande waterdruk in het systeem voldoet tot en met de derde verdieping. Omdat het onderzochte gebouw echter 4 verdiepingen telt, dient de hydrofoorinstallatie ingezet te worden. Verbeterpunten voor de koudwaterinstallatie zijn het inzetten van spoelonderbrekers voor de WC's. Tevens dient een hydrofoorinstallatie ingezet te worden om de waterdruk in het koudwatersysteem op de juiste hoogte te brengen.

figuur 4-3 Voorbeeld dynamisch tekstblok

In deze paragraaf wordt beschreven hoe een rapportgenerator, met behulp van de in het vorige hoofdstuk beschreven technieken, zo'n tekst kan genereren. Als de exacte tekst in iedere rapportage voorkomt, bijvoorbeeld omdat het een standaardinleiding is, is het verstandig om de tekst als geheel op te nemen in het rapportmodel, zoals besproken in paragraaf 4.2. Dit geldt ook voor de figuren en tabellen. De rapportgenerator zal de inhoud van figuren en tabellen niet genereren. Deze zullen vooraf gegeneerd zijn en door de rapportgenerator eventueel in het rapport geplaatst worden.

De tekst kan ook in het rapport geplaatst worden met behulp van de op regels gebaseerde of Case Based Reasoning techniek, eventueel in combinatie met sjablonen. Bij de op regels gebaseerde techniek moet dan allereerst voor iedere zin of combinatie van zinnen bepaald worden onder welke voorwaarden de zin in het uiteindelijke rapport geplaatst zal worden. De getallen kunnen vervolgens ingevuld worden met behulp van sjablonen, zoals in de derde zin van het voorbeeld:

De wc's, die niet zijn voorzien van spoelonderbrekers, hebben een waterreservoir van 20 liter.

Deze zin kan gegenereerd worden op basis van onderstaand sjabloon:

De WC's, die **wel_niet** zijn voorzien van spoelonderbrekers, hebben een waterreservoir van **inhoud_waterreservoir** liter.

figuur 4-4 Voorbeeld sjabloon

Voor de invulling van het sjabloon is aan **wel_niet** de waarde *wel* toegekend en aan **inhoud_waterreservoir** *20*. Deze sjablonentechniek kan ook toegepast worden voor de andere zinnen van het voorbeeld.

Voor de NLG techniek is meer kennis nodig dan alleen te weten dat de inhoud van het waterreservoir 20 liter is. De NLG techniek maakt gebruik van een gedetailleerde beschrijving van de werkelijkheid waaronder de relatie tussen de verschillende objecten en hun eigenschappen. De relatie tussen WC, spoelonderbrekers en waterreservoir moet dus bekend zijn, maar ook dat een waterreservoir de eigenschap *inhoud* heeft. Dit vereist een uitgebreid onderzoek naar de kennis uit het kennisdomein en zal voor het genereren van grote rapporten met veelal dezelfde teksten niet lonend zijn. Bovendien is er een uitgebreide kennis van de grammaticaregels benodigd om zinnen te genereren op basis van de kennis uit het kennisdomein.

4.4 Invulling (2)

In de vorige paragraaf is behandeld hoe de tekst van een rapport ingevuld kan worden met behulp van de NLG techniek en met behulp van een op regels gebaseerd systeem op basis van sjablonen. Gesteld werd dat de pure NLG techniek aantal voordelen ten opzichte van de sjablonen techniek heeft, maar deze wegen niet op tegen het grote nadeel. Het initiële vastleggen van de benodigde kennis van de expert, maar ook van de grammatica is zeer complex en tijdrovend. Dit is zelfs zo complex dat er volgens Dale in [4] nog geen voorbeelden van commerciële systemen zijn waarin deze pure NLG techniek is toegepast. Toch zijn een aantal van de ideeën binnen de pure NLG architectuur goed te gebruiken bij het maken van rapporten. De zinnenplanner van de pure NLG techniek kan gecombineerd worden met de sjablonen van paragraaf 3.4.3. Deze combinatie is uit te leggen aan de hand van het voorbeeld van figuur 4-3. In dit voorbeeld staat onder andere de volgende zin:

Het koudwatersysteem bestaat uit tapkranen en spoelknoppen.

figuur 4-5 Voorbeeldzin combinatie

In de vorige paragraaf werd een sjabloon gebruikt om een soortgelijke zin te genereren. Er is echter ook een andere manier om tot de zin van dit voorbeeld te komen. Allereerst kunnen de volgende twee sjablonen gebruikt worden om de zinnen van figuur 4-8 te genereren, ervan uitgaande dat **tapkranen** en **spoelknoppen** bestaan en respectievelijk de waarden *tapkranen* en *spoelknoppen* bevatten:

Het koudwatersysteem bestaat uit **tapkranen**.

figuur 4-6 Sjabloon 1

Het koudwatersysteem bestaat uit **spoelknoppen**.

figuur 4-7 Sjabloon 2

Het koudwatersysteem bestaat uit tapkranen.
Het koudwatersysteem bestaat uit spoelknoppen.

figuur 4-8 Resultaat van de sjablonen

Er wordt voor de twee sjablonen geen gebruik meer gemaakt van criteria, deze zijn namelijk impliciet af te leiden uit de zin zelf. Mocht **tapkranen** bijvoorbeeld niet voorkomen in de invoerdata dan hoeft de hele zin van figuur 4-6 niet in het resultaat geplaatst worden. Als **tapkranen** en **spoelknoppen** wel bestaan worden twee zinnen gegenereerd die weliswaar correct Nederlands zijn, maar niet echt een vloeiend leesbare combinatie vormen. Vanaf dat moment komt de zinnenplanner in actie. Deze twee zinnen kunnen

samen gevoegd worden met de in die paragraaf beschreven techniek van de zinnenplanner. Het resultaat zal de zin van figuur 4-5 zijn.

De hierboven beschreven combinatie van sjablonen en zinnenplanner creëert niet alleen duidelijk leesbare zinnen, maar zorgt ook voor een eenvoudige opslag van de te genereren rapporten. Nu hoeft niet voor ieder sjabloon criteria opgesteld te worden, maar kan het systeem deze zelfstandig afleiden. Dit wil echter niet zeggen dat er helemaal geen criteria meer gebruikt hoeven te worden. Voor sommige zinnen zal het nog steeds nodig zijn om criteria te gebruiken die door de expert zijn opgesteld.

Er is echter nog een mogelijke combinatie tussen de twee NLG technieken te bedenken die zich speciaal richt op verzamelingen. De twee genoemde sjablonen kunnen herschreven worden tot één sjabloon:

Het koudwatersysteem bestaat uit `onderdelen_koudwatersysteem`.

figuur 4-9 Verzamelingen Sjabloon

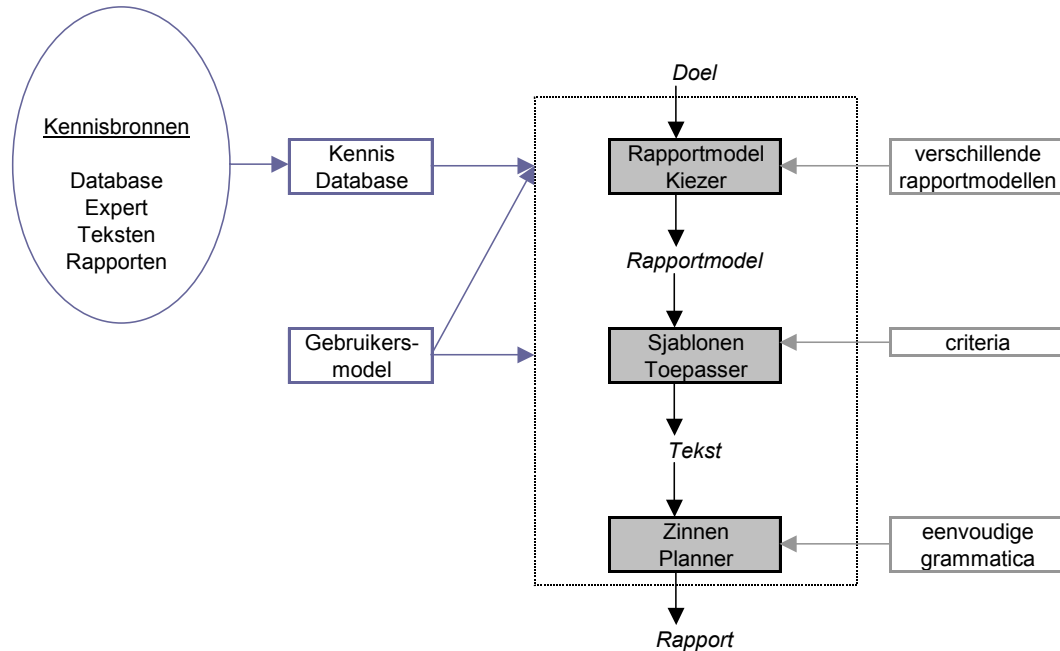
Er wordt hier gebruik gemaakt van de samengestelde variabele `onderdelen_koudwatersysteem`. In de voorafgaande hoofdstukken werd slechts gebruik gemaakt van enkelvoudige variabelen zoals een getal, een woord of een (deel van een) zin. De uitbreiding met samengestelde variabelen is vooral nuttig bij het schrijven van grote opsommingen. Ten eerste omdat het aantal variabelen terug gebracht wordt, in dit geval van twee (`spoelknoppen` en `tapkranen`) naar één (`onderdelen_koudwatersysteem`). Ten tweede omdat er nog maar één sjabloon geschreven hoeft te worden. Dit voorkomt de combinatorische explosie zoals gesteld in paragraaf 3.2.3.

4.5 Architectuur

De in dit hoofdstuk beschreven architectuur van een rapportgenerator voldoet aan alle in de inleiding gestelde eisen, te weten:

- Het systeem moet de inhoud van het rapport op hoofdlijnen kunnen bepalen, dus de indeling in hoofdstukken, paragrafen en standaardteksten kunnen realiseren. Dit is gerealiseerd met behulp van de rapportmodellen van paragraaf 4.2.
- Het systeem kan het rapport met tekstblokken, figuren en tabellen invullen aan de hand van gestelde criteria. Hiervoor kan wederom het rapportmodel van paragraaf 4.2 gebruikt worden.
- Het systeem moet teksten kunnen samenstellen uit losse zinnen, waarbij iedere zin afhankelijk is van bepaalde criteria en gebruik maakt van bepaalde variabelen. Dit kan met sjablonen met of zonder criteria van respectievelijk paragraaf 4.3 en 4.4.

Toch is het niet zo dat met deze architectuur een perfect rapport gegenereerd kan worden dat exact gelijk is aan dat van een expert. Enerzijds is het zo dat afzonderlijke experts al verschillende rapporten zullen schrijven en het automatisch gemaakte rapport niet met één ideaal rapport vergeleken kan worden. Anderzijds geldt dat een experts vaak voor het systeem onbeschikbare data gebruiken. Dit laatste probleem is zeer moeilijk op te lossen omdat een computer onmogelijk alle kennis van een expert kan bevatten. Dat wil niet zeggen dat een door de computer gemaakt rapport onbruikbaar is. In veel gevallen zal een gegenereerd rapport al voldoende zijn voor de wensen en eisen van de gebruiker. In andere gevallen zal de expert nog handmatig aanpassingen aan het rapport moeten maken. Zelfs in gevallen waar een volledig handmatig geschreven rapport gewenst is, kan een gegenereerd rapport van nut zijn, omdat de expert zich hierop kan baseren bij het schrijven van het gewenste rapport. Schematisch kunnen de in dit en het vorige hoofdstuk beschreven technieken worden samengevat tot een architectuur voor een rapportgenerator.



Figuur 4-10 Architectuur Overzicht

Deze architectuur lijkt op de in figuur 3-10 gepresenteerde NLG Architectuur van Reiter en Dale. Toch zijn er grote verschillen tussen de twee architecturen. Allereerst wordt met behulp van de Rapportmodel Kiezer een rapportmodel gekozen aan de hand waarvan het uiteindelijke rapport gemaakt zal worden. Daarna komt de Criteria Toepasser in actie die op basis van vooraf opgestelde criteria tekstblokken overneemt en zinnen invult. Ten slotte maakt de Zinnen Planner lopende zinnen van de door de Sjablonen Toepasser geproduceerde teksten. De Zinnen Planner heeft een lastige taak uit te voeren als hij iedere combinatie van zinnen moet proberen samen te voegen met alle mogelijke aggregatiemethoden. Vandaar dat de Zinnen Planner in deze architectuur slechts twee of meer opeenvolgende zinnen kan samenvoegen en alleen als deze zinnen gebruik ontstaan zijn aan de hand van een sjabloon. Dit wil zeggen dat vaste delen van de tekst van het rapport ongemoeid worden gelaten door de Zinnen Planner.

De in dit hoofdstuk besproken architectuur is heel algemeen en richt zich niet op een specifiek rapport. In het volgende hoofdstuk wordt gekeken naar de EPA rapportage en hoe met behulp van deze architectuur de EPA rapporten gemaakt kunnen worden. Daar wordt ook besproken hoe de door de rapportgenerator te gebruiken kennis verkregen en vastgelegd kan worden.

Hoofdstuk 5

Geautomatiseerd EPA Systeem

De in het vorige hoofdstuk behandelde architectuur heeft als doel het opslaan en gebruiken van de kennis van een expert bij het schrijven van een rapport. Dit hoofdstuk richt zich op het gebruik van deze rapportgenerator in de geautomatiseerde uitvoering van een EPA. Daarom wordt allereerst bekeken aan welke eisen het geautomatiseerde EPA systeem, inclusief de rapportgenerator, moet voldoen om ingezet te kunnen worden bij de uitvoering van een EPA. Vervolgens wordt schematisch weergegeven uit welke componenten het geautomatiseerde EPA systeem bestaat en in welke van de drie in hoofdstuk 2 beschreven fasen van een EPA uitvoering deze componenten ingezet zullen worden. In dit hoofdstuk wordt vooral de samenhang tussen de verschillende componenten geschetst en de ervoor te gebruiken informatie. In het volgende hoofdstuk wordt dieper ingegaan op het software ontwerp voor dit EPA systeem.

5.1 Waarom automatiseren?

Om een rapportgenerator, zoals in het vorige hoofdstuk voorgesteld, succesvol in te zetten, zal duidelijk moeten zijn welke informatie de rapportgenerator daarbij kan gebruiken. Voor het genereren van EPA rapporten zal onderzocht moeten worden welke informatie er van een ondernemer benodigd is en hoe deze ingewonnen kan worden.

Een geautomatiseerde versie van het EPA traject heeft de volgende voordelen ten opzichte van een niet-geautomatiseerde versie:

- Een ondernemer kan zelf via internet bepalen of het noodzakelijk of nuttig is een EPA te laten uitvoeren.
- Een EPA adviseur kan met hetzelfde systeem snel zien met wat voor soort onderneming hij te maken heeft.
- Een EPA adviseur kan een EPA vragenlijst laten maken door het systeem. Hij hoeft dus niet zelf een vragenlijst te maken.
- Een EPA adviseur kan een rapport op maat door het systeem laten maken dat (na enige aanpassing) direct presenteerbaar is voor de ondernemer.

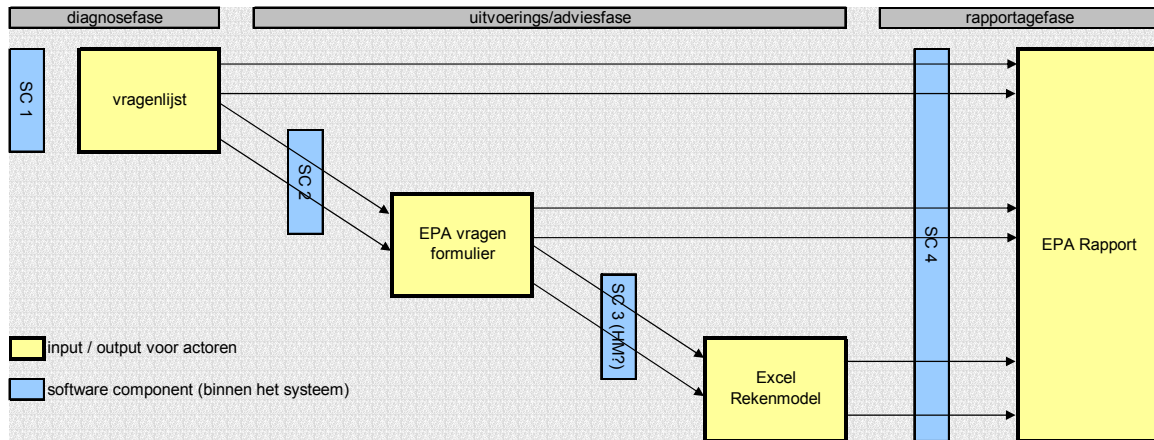
5.2 Wat automatiseren?

Om het geautomatiseerde EPA systeem met succes binnen het EPA traject te kunnen gebruiken, moeten een aantal eisen aan het prototype gesteld worden. Hieronder staat het programma van eisen opgesomd:

- Het systeem moet via internet bereikbaar zijn voor zowel de ondernemer, de adviseur, als de beheerder; zo kunnen de ingevoerde gegevens eenvoudig centraal opgeslagen worden.
- Het systeem kan aan de hand van enkele eenvoudige vragen aan de ondernemer bepalen of:
 - de ondernemer wettelijk verplicht is een EPA te laten uitvoeren,
 - de ondernemer zich op een natuurlijk moment bevindt om een EPA te laten uitvoeren,
 - het niet nodig is voor de ondernemer om een EPA te laten uitvoeren.
- Het systeem moet conclusies trekken over de toestand waarin een onderneming zich bevindt om zo te kunnen bepalen wat de inhoud van het EPA rapport dat aan ondernemer aangeboden wordt, moet worden.
- Het systeem kan aan de hand van de antwoorden op de bovengenoemde vragen de vragenlijsten van het EPA onderzoek, dus de vragen die door de adviseur op locatie bij de ondernemer beantwoord moeten worden, bepalen.
- Het systeem kan berekeningen met behulp van een rekenmodel uitvoeren.
- Het systeem kan aan de hand van ingevulde EPA vragenlijsten en de uitkomsten van het rekenmodel een EPA rapport op maat genereren voor de onderneming. Dit rapport is niet noodzakelijkerwijs volledig en foutloos en dient daarom door de EPA adviseur waar nodig aangepast en aangevuld te worden.

Om een rapportgenerator, zoals in het vorige hoofdstuk voorgesteld, succesvol in te zetten, zal duidelijk moeten zijn welke informatie de rapportgenerator daarbij kan gebruiken. Voor het genereren van EPA rapporten zal onderzocht moeten worden welke informatie er van een ondernemer benodigd is en hoe deze ingewonnen kan worden. De procedure hiervoor wordt gebaseerd op de normale uitvoering van een EPA, maar met gebruik van het internet. In hoofdstuk 2 werd een schematisch overzicht gegeven van de verschillende stappen binnen een normaal EPA traject. Een geautomatiseerde versie van het EPA traject zal grotendeels dezelfde stappen bevatten als de in het vorige hoofdstuk genoemde stappen van een normaal EPA traject. De geautomatiseerde uitvoering van het EPA traject kan dan ook onderverdeeld worden in drie belangrijke fasen. Allereerst de diagnosefase waarin wordt bekeken met wat voor soort onderneming het systeem te maken heeft. Daarna volgen de adviesfase en de uitvoeringsfase, dit zijn wat betreft tijdsduur de langste fasen omdat hierin het EPA onderzoek bij de ondernemer wordt uitgevoerd. Het systeem levert in deze fase de EPA vragenlijsten aan, waarmee de adviseur het EPA onderzoek uitvoert. Als laatste onderdeel van het EPA traject volgt de rapportagefase. Hierin wordt aan de hand van de gegevens uit de voorafgaande fasen het rapport voor de ondernemer samengesteld.

Het schema van figuur 2-1 kan worden vertaald naar een systeemoverzicht waarin een viertal software componenten te vinden zijn.



figuur 5-1 Systemoverzicht EPA

In het bovenste deel van dit figuur staan wederom de drie fasen van een EPA onderzoek beschreven, zodat te zien is in welke fase een bepaalde software component gebruikt zal gaan worden. Deze drie fasen vormen de leidraad voor de behandeling in dit hoofdstuk van de verschillende componenten.

In het kort hebben de verschillende softwarecomponenten de volgende functionaliteit:

1. SC 1 zorgt voor het presenteren van een korte vragenlijst aan de ondernemer en bepaalt daarna of het verstandig is dat de ondernemer een uitgebreid EPA onderzoek laat uitvoeren door een EPA adviseur.
2. SC 2 maakt met behulp van de gegevens van de ondernemer een EPA vragenlijst. Dit is een uitgebreid onderzoeksformulier voor de EPA adviseur. Aan de hand van dit formulier voert hij zijn onderzoek bij de ondernemer op locatie uit.
3. SC 3 heeft als invoer de onderzoeksresultaten van SC 2.
4. SC 4 gebruikt de resultaten van de vragenlijst, het ingevulde EPA formulier en de uitkomsten van het rekenmodel om hieruit een EPA rapport samen te stellen.

5.3 Procesbeschrijving geautomatiseerde EPA

In de vorige paragraaf is behandeld wat er in de uitvoering van een EPA geautomatiseerd kan worden en welke software componenten daarbij gebruikt dienen te worden. Dit systeem biedt ondersteuning voor de bij een uitvoering van een EPA benodigde stappen, zoals blijkt uit de onderstaande procesbeschrijving. Deze procesbeschrijving bevat grotendeels dezelfde stappen als de in paragraaf 2.3 beschreven stappen.

- Stap 1 Vragen stellen ter identificatie van de ondernemer
 - In: De startvragenlijst
 - Uit: Antwoorden
- Stap 2 Opstellen klantprofiel
 - In: Antwoorden stap 1
 - Uit: Klantprofiel
- Stap 3 Bepaling behoefte klant
 - In: Klantprofiel
 - Uit: Passend adviessoort
 - Uit: Passende vragenlijst
 - Uit: Passend presentatiesoort
- Stap 4 Uitvoeren EPA doorlichting
 - In: Vragenlijst van stap 3
 - Uit: Antwoorden op vragen van vragenlijst
- Stap 5 Berekeningen e-functies (gebruik van rekenmodel)
 - In: Klantprofiel
 - In: Antwoorden uit stap 4
 - Uit: Sluitende energiebalans
 - Uit: Berekende besparingen
- Stap 6 Bepalen passende rapportvorm
 - In: Klantprofiel
 - In: Antwoorden uit stap 4
 - Output: Rapportvorm
- Stap 7 Genereren EPA rapport
 - In: Uitkomsten van berekeningen e-functies
 - In: Geselecteerde rapportmodel van stap 6
 - Uit: Advies in rapportvorm

E-functies zijn rekenmethoden aan de energiehuishouding van een bedrijf en vormen als geheel het rekenmodel. Met een e-functie kan bijvoorbeeld het gemiddeld energieverbruik van een onderneming per maand berekend worden.

5.4 Diagnosevragen

De eerste fase binnen het EPA traject die hier behandeld wordt, is de diagnosefase. Binnen de niet-geautomatiseerde variant van het EPA traject bestaat dit uit een (telefoon-) gesprek tussen een ondernemer en een EPA adviseur. Het geautomatiseerde EPA systeem biedt de ondernemer de mogelijkheid deze fase te doorlopen via het web.

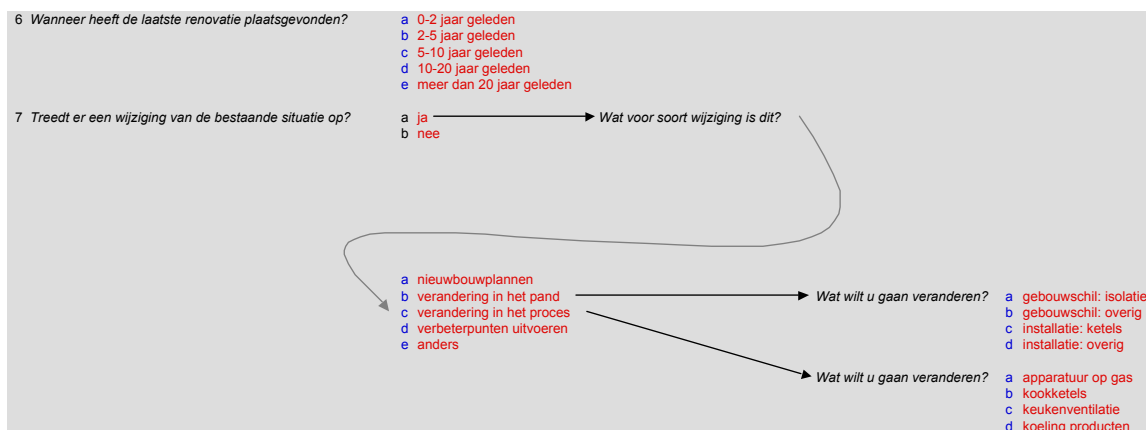
De vragen die in dit gedeelte van het EPA traject aan de ondernemer gesteld worden, zijn eenvoudig van aard. Dit betekent dat er alleen vragen gesteld worden waar de ondernemer direct of na kort zoekwerk antwoord op kan geven.

Voorbeelden van vragen die aan de ondernemer gesteld kunnen worden, zijn:

- Als wat voor soort onderneming is uw onderneming te typeren?
- Wat is het elektraverbruik van uw onderneming (in kWh)?
- Wat is het bouwjaar van uw bedrijfspand?

De genoemde vragen worden in de niet-geautomatiseerde EPA uitvoering door de EPA adviseur gesteld aan de ondernemer. Deze functionaliteit kan overgenomen worden door het EPA prototype met een vragenlijst in de vorm van een webpagina. Daarvoor is onderzocht welke vragen aan de ondernemer gesteld worden. Van belang is ook in welke volgorde deze vragen gesteld worden om het stellen van overbodige vragen te voorkomen. Hiervoor is een vragenschema opgesteld dat als basis van de eerste module van de geautomatiseerde EPA zal dienen. In dit vragenschema, waarvan een gedeelte in figuur 5-2 te zien is, zijn de volgende zaken opgenomen:

- De te stellen vragen.
- De mogelijke antwoorden.
- De volgorde van de vragen.
- De relaties tussen de vragen.



figuur 5-2 Gedeelte van vragenschema

Er wordt in het vragenschema slechts uitgegaan van meerkeuzevragen. Daardoor zijn enerzijds de vragen voor de ondernemer eenvoudiger te beantwoorden. Anderzijds kan het systeem daarna beter de antwoorden verwerken, zoals blijkt in paragraaf 5.5. Een verdere vereenvoudiging voor de ondernemer is om meerdere vragen op een enkele pagina te stellen, in plaats van steeds één vraag. Alle mogelijke vragen tegelijk op één scherm aanbieden is echter weer verwarrend voor de ondernemer. Bovendien is dit geen verstandige strategie omdat er dan onnodige vragen worden gesteld. Daarom is er gekozen voor een methode waarbij er nieuwe vragen op het scherm bijkomen afhankelijk van de door de ondernemer gekozen antwoorden.

5.5 Klantprofiel

Aan de hand van de uitkomsten op de vragen uit de vorige paragraaf moet het systeem een verzameling van vooraf gedefinieerde conclusies kunnen trekken waarop de onderneming in een bepaalde categorie ingedeeld kan worden.

Een voorbeeld van een conclusie die het systeem kan trekken is bijvoorbeeld dat de ondernemer wettelijke zorg- en onderzoeksplicht heeft omdat het elektraverbruik van zijn onderneming hoger is dan 25.000 kWh. Dit betekent dat hij wettelijk verplicht is om een EPA onderzoek te laten uitvoeren en bepaalde maatregelen zoals voorgesteld in zo'n rapport zal moeten nemen. Deze conclusie heeft grote gevolgen voor het uiteindelijke rapport dat de ondernemer gepresenteerd krijgt; een ondernemer die niet wettelijk verplicht is energemaatregelen te laten uitvoeren zal juist door het rapport overgehaald moeten dit wel te doen, terwijl een andere ondernemer al door de letteren der wet is overtuigd (of overtuigd zou moeten zijn)

Ook kan het systeem de aanneme doen dat de ondernemer bereid is tot besparingsmaatregelen met een terugverdientijd dan meer dan 5 jaar omdat hij aangeduid had dat de besparingsmaatregelen geld mogen kosten en niet van plan was op korte termijn te stoppen met de onderneming.

Deze verzameling van veronderstellingen en conclusies wordt het klantprofiel genoemd en is een belangrijke bron van informatie over de onderneming. Het is bovendien de basis waarop de beslissing om door te gaan naar de adviesfase wordt gebaseerd.

In de onderstaande tabel is een gedeelte van een ingevuld klantprofiel te vinden.

Eigenschap	Waarde
Branche	Horeca
Sector	Café / Bar
Motivatie	Zeer
Natuurlijk moment	Ja
Laatste renovatie bedrijfspand	1983
Wettelijke verplichting voor maatregelen	Nee

tabel 5-1 Gedeelte van een klantprofiel

5.6 Vragenlijsten

In de diagnosefase werd voornamelijk algemene informatie van de onderneming gevraagd. Deze informatie was allereerst nodig om te bepalen of het wel noodzakelijk of nuttig was voor de ondernemer om een EPA te laten uitvoeren voor zijn onderneming. De hierbij opgevraagde informatie zoals verzameld in het klantprofiel kan in de uitvoeringsfase gebruikt worden om te bepalen welke gedetailleerde informatie er van de onderneming benodigd is om te komen tot een EPA rapport.

De benodigde informatie zal met behulp van vragenlijsten ingewonnen worden. Gedeelten van deze vragenlijsten zullen echter niet van toepassing zijn voor een specifieke onderneming. Zo zal het nooit nuttig zijn om aan de houder van een ijscoke vragen te stellen over de zwembadpompinstallatie van zijn onderneming.

Het systeem zal daarom op basis van het klantprofiel een vragenlijst genereren waarin alleen naar mogelijk van toepassing zijnde aspecten van de onderneming gevraagd wordt. Dit betekent niet dat de vragenlijst voor iedere onderneming volledig ingevuld kan worden, zo heeft niet iedere ijscoke een frituur. De kans is echter zo klein mogelijk dat er delen van vragenlijsten zijn die nooit ingevuld zullen worden.

Op de volgende pagina is een gedeelte van een mogelijke vragenlijst weergegeven.

EPA - KENTALLEN VRAGENLIJST VOOR KHN-LEDEN

A. Algemeen

Relatienummer KHN

Naam Cafe / Bar

Straat en nummer van bedrijfspand

Postcode van bedrijfspand

Vestigingsplaats van bedrijfspand

Postadres

Naam directeur/ eigenaar

Contactpersoon , naam & functie

Rechtsvorm BV NV VOF Zelfstandige

Telefoonnummer

Faxnummer

E-mail adres

Klantnummer

Totale Bruto Vloer Oppervlak (m²)
of begane grond opmeten : Lengte (m) Breedte (m)

Gemiddelde hoogte per verdieping

Aantal verdiepingen

Aantal medewerkers

Aantal klanten per jaar (schatting)

Omzetcijfer (€ per jaar)

Typering onderneming in winkelcentrum onder woonhuis
 café zelfstandig

Bouwjaar bedrijfspand

Weet niet, indicatie: < 1950 1950-1990 > 1990

Wanneer was de laatste renovatie?

Hoe lang bent u in het pand gevestigd?

Is het pand in bezit of gehuurd?

Openingstijden

	maandag	dinsdag	woensdag	donderdag	vrijdag	zaterdag	zondag
van	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
tot	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

1 van 9

figuur 5-3 Vragenlijst

5.7 EPA rekenmodel

Het EPA rekenmodel is een module die op basis van invoergegevens berekeningen uitvoert aan de energiehuishouding van een onderneming. Het rekenmodel gebruikt de informatie zoals opgevraagd met de vragenlijsten van de vorige paragraaf en voert op basis hiervan berekeningen uit aan de energiehuishouding van een onderneming. Het resultaat van deze berekeningen zijn opsommingen, getallen, tabellen en figuren die gebruikt kunnen worden voor plaatsing in een EPA rapport.

In het rekenmodel worden o.a. berekeningen uitgevoerd aan:

- Ventilatie van het gebouw;
- Be- en ontvochtiging van het gebouw;
- Energiebesparingsmaatregelen;
- Verlichting;
- Gebouwschil;
- Warmteopwekking;
- Ruimtekoeling;
- Gasgestookte en elektrische keukenapparatuur;
- Warmwaterhuishouding;
- Reinigingsapparatuur;
- Enz....

5.7.1 Kengetallen

Indien er voor één van de genoemde berekeningen geen invoergegevens beschikbaar zijn, dus omdat er niets is ingevuld in de vragenlijst, of omdat het onderwerp zelfs niet op vragenlijst stond, wordt de berekening niet uitgevoerd. Zo zal er voor een ijscokar geen informatie beschikbaar zijn over de ventilatie, omdat deze hierin niet aanwezig is. Het rekenmodel levert dan geen resultaten op over de ventilatie.

Als er voor een bepaalde berekening gegevens ontbreken, zal het rekenmodel kengetallen gebruiken ter vervanging van de ontbrekende gegevens. Een kengetal is een verhoudingscijfer dat een verband aangeeft tussen verschillende grootheden, met de bedoeling prestaties en ontwikkelingen zichtbaar te maken en het systematisch nadenken over het gericht verbeteren van de uitvoering te stimuleren. Voor het rekenmodel is een verzameling aannamen en kengetallen aangelegd waarvan in de berekeningen gebruik gemaakt kan worden. Voor berekeningen aan de gebouwschil is bijvoorbeeld de gemiddelde temperatuur in de onderzochte ruimte nodig. Indien deze niet bekend is, wordt de gemiddelde temperatuur van gelijksoortige ondernemingen toegepast. Voor de gebouwschil kunnen bijvoorbeeld aannamen gedaan worden over de u -waarde⁷ van bijvoorbeeld het dak. Deze u -waarde is afhankelijk van het gebruikte isolatiemateriaal. Indien dit niet bekend is, kan hiervoor de gemiddelde u -waarde van de daken van gebouwen uit dezelfde bouwperiode genomen worden. De gemiddelde waarden worden met ieder ingevuld antwoord in de vragenlijst verder bijgewerkt, zodat het opgeslagen gemiddelde (in theorie althans) steeds dichterbij het echte gemiddelde komt.

⁷ u -waarde: Geleidingswaarde (van bijvoorbeeld raam of deur) Dit is wederkerig aan de isolatiewaarde.

5.7.2 Meetwaarden

Het rekenmodel gebruikt naast (continue aangepaste) aannamen en kencijfers ook tabellen met constanten en niet-veranderende meetwaarden, bijvoorbeeld de temperatuurgegevens van de afgelopen 10 jaar. Eén van de gebruikte tabellen, met de gemiddelde dagtemperatuur van vijf meetstations in Nederland is hieronder gegeven.

Periode	De Kooy	Eelde	De Bilt	Twente	Vlissingen
jan-91	3,6	2,6	3,2	2,9	4,1
feb-91	-0,6	-1,1	-0,8	-0,9	0,5
mrt-91	7,5	7,6	8,8	8,3	8,1
apr-91	8,1	7,7	8,5	8,1	8,8
mei-91	9,4	9,1	10	9,7	9,9
jun-91	12,3	12,1	12,7	12,6	13,1
jul-91	18,1	18,4	19	18,8	18,2
aug-91	17,6	17,1	18	17,3	18,7
sep-91	15,5	14,4	15	14,7	16,8
okt-91	10,7	9,2	10,2	9,4	12,2
nov-91	6,6	5	5,3	5	6,9
dec-91	5,1	3,2	3,9	3,2	4,7
jan-92	4,1	2,8	2,7	2,2	3,6
feb-92	5,2	4,6	4,9	4,7	5,2
mrt-92	6,4	6	6,9	6,4	7,2
apr-92	8,2	7,9	8,7	8,3	9,2
mei-92	14,7	14,7	15,6	15,4	15,1
jun-92	16,8	17,1	17,2	17	16,8
jul-92	17,7	17,5	18,3	18,2	18,4
aug-92	17,2	17	17,8	18	18,2
sep-92	14,7	13,8	14,6	14,3	15,4
okt-92	8,9	7,2	8	7,3	9,7
nov-92	8,2	6,9	8	7,4	8,7
dec-92	4,3	3	3,7	3,3	4,9

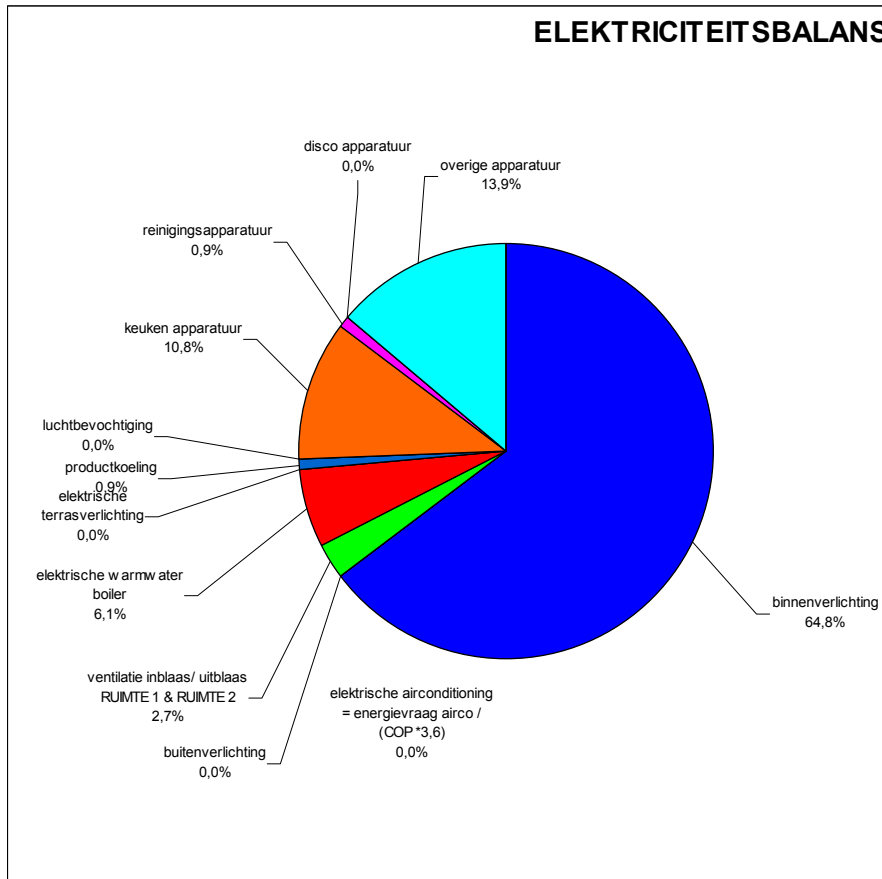
tabel 5-2 Gedeelte van meetwaardentabel

5.7.3 Besparingsmaatregelen

Het rekenmodel selecteert ook mogelijke energiebesparingsmaatregelen, o.a. op basis van de terugverdientijd, en berekent het effect hiervan op de energiehuishouding van de onderneming. Een voorbeeld van een energiebesparingsmaatregel is bijvoorbeeld het vervangen van de huidige frituurpan door een gasgestookte frituurpan. Van ieder van deze maatregelen is de terugverdientijd en de mogelijke financiële besparing bekend. Alle maatregelen kunnen onafhankelijk van elkaar uitgevoerd worden, maar het besparingseffect van opeenvolgende maatregelen zal steeds minder worden.

5.7.4 Uitkomsten

Het rekenmodel levert uitkomsten in de vorm van tabellen en figuren. Een van die figuren is de elektriciteitsbalans van de onderneming.



figuur 5-4 Elektriciteitsbalans voorbeeld

Energie Prestatie Advies

In de tabellen staan de uitkomsten van berekeningen en de vastgelegde meetwaarden.

Type	Verbruik
elektriciteit	49543 kWh
gas	22294 m3
warmte	0 GJ
water	150 m3

figuur 5-5 Voorbeeld energieverbruiktabel (gedeeltelijk)

Besparingsmaatregel	Al toegepast
Schakelklok op Ventilatie	ja
Hoog Rendement CV Ketel	nee
CV-leiding Isolatie	nee
Hoog Frequentie Verlichting	ja

figuur 5-6 Voorbeeld toegepaste maatregelentabel (gedeeltelijk)

Besparingsmaatregel	Volgorde van invoering
Isoleren dak	1
Gasfornuis i.p.v. elektrisch fornuis	2
Doorstroombegrenzer	3
Waterbesparende toiletten	4

figuur 5-7 Voorbeeld voorgestelde maatregelentabel (gedeeltelijk)

5.8 Rapportgenerator

Voor het genereren van EPA rapporten in de rapportagefase kan er gebruik gemaakt worden van de tekst- en rapportgeneratietechnieken zoals besproken in de twee voorgaande hoofdstukken. De in hoofdstuk 4 voorgestelde techniek gaat uit van meerdere rapportmodellen waaruit een keuze gemaakt moet worden. Zo'n rapportmodel dient als basis voor het te genereren rapport en zal ingevuld worden met teksten, figuren en tabellen.

In deze paragraaf wordt de rapportagetechniek met behulp van rapportmodellen verder uitgewerkt en wordt beschreven hoe zo'n rapportmodel in software gerepresenteerd kan worden.

Er zijn vele mogelijkheden te bedenken waarmee het rapportmodel opgeslagen kan worden. Er kan gebruik worden gemaakt van het formaat waarmee het uiteindelijke rapport gepresenteerd wordt. Als het uiteindelijke rapport bijvoorbeeld gerealiseerd zal worden in een Microsoft Word document kan dit formaat ook gebruikt worden voor het rapportmodel. Het nadeel van het gebruik van een rapportmodel in het uitvoerformaat is dat als er na verloop van tijd gekozen wordt voor een ander uitvoerformaat, bijvoorbeeld een HTML bestand in plaats van een Microsoft Word document, dat het rapportmodel aangepast of zelfs opnieuw gemaakt moet worden, afgezien conversiemogelijkheden van het ene naar het andere formaat. Beter is om gebruik te maken van een uitvoeronafhankelijk formaat. Dit kan bijvoorbeeld door het rapportmodel op te slaan in een relationele database, een case base, een eigen formaat of in XML (eXtensible Markup Language). Met elk van deze formaten kan op een gestructureerde manier het rapportmodel onafhankelijk van het uitvoerformaat opgeslagen worden.

XML is een relatief jonge standaard voor het World Wide Web. Zij is ontwikkeld en gestandaardiseerd door het W3 consortium, de organisatie die verantwoordelijk is voor standaarden op het World Wide Web zoals HTML en HTTP. Met behulp van XML is het mogelijk om in tekstbestanden met zogenaamde tags tekstfragmenten te markeren. Een tag⁸ heeft een naam en om het onderscheid te maken met de tekst in het tekstbestand zelf, begint een tag met het teken < en eindigt met het teken >. Er zijn twee soorten tags: begintags en eindtags. Eindtags onderscheiden zich van begintags door het teken /. Tussen de tags staat het tekstfragment dat wordt gemarkeerd.

```
<auteur>C. Boer</auteur>
```

figuur 5-8 XML tag voorbeeld

XML is dus een verzameling regels die voorschrijft hoe tags eruit zien en hoe ze mogen worden gebruikt om tekstfragmenten in een tekstbestand te markeren. De namen van de tags zijn niet vastgelegd. De belangrijkste regels zijn hieronder opgesomd:

- Tags moeten beginnen met een < en eindigen met een >
- Eindtags beginnen met </

⁸ Tag (engels): label, etiket

- Als er een begintag is, moet er een eindtag zijn met dezelfde naam.
- In tagnamen bestaat er verschil tussen hoofdletters en kleine letters (case-sensitive)
- Het tekstbestand moet in zijn geheel door een begin- en eindtag omvat worden.
- Als een fragment wordt gemarkeerd binnen een ander fragment, moet het binnenste fragment eerst zijn afgesloten, alvorens het buitenste fragment mag worden afgesloten.
- Als er niets tussen een begin- en eindtag staat, mag deze afgekort worden tot <tagnaam />.

Hier zal uitgegaan worden van een rapportmodel in XML om de volgende redenen:

- XML bestanden zijn met ieder tekstverwerkingsprogramma en ontwikkelomgeving te maken.
- De structuur van het rapportmodel kan eenvoudiger worden vastgelegd in XML formaat dan in bijvoorbeeld een relationele database.
- De structuur van het rapportmodel kan in XML eenvoudig uitgebreid of aangepast worden.
- Voor ieder mogelijk rapportmodel kan een afzonderlijk XML bestand gemaakt worden; dit komt de beheerbaarheid ten goede.

Het maken van een rapportmodel in XML bestaat uit twee stappen. Allereerst moet met behulp van een DTD (Document Type Declaration; beschrijving van de structuur van een XML document) de structuur van het XML rapportmodel vastgelegd worden. Een DTD definieert de structuur en mogelijk inhoud van een XML document. In figuur 5-9 staat een eenvoudig DTD van een mogelijk rapportmodel.

```
<!ELEMENT rapportmodel (criterium?, hoofdstuk*)>
<!ELEMENT criterium (#PCDATA)>
<!ELEMENT hoofdstuk (koptekst, criterium?, paragraaf*)>
<!ELEMENT paragraaf (koptekst, criterium?)>
<!ELEMENT koptekst (#PCDATA)>
```

figuur 5-9 Rapportmodel DTD

Deze DTD duidt aan dat het rapportmodel één of geen criterium bevat en uit nul of meerdere hoofdstukken bestaat. Een hoofdstuk bestaat uit een koptekst en 0 of meerdere paragrafen. Ook een paragraaf kan een criterium bevatten. Een paragraaf kan ook een criterium bevatten.

Een criterium bepaalt of een hoofdstuk of paragraaf in het uiteindelijke rapport geplaatst moet worden. Dit kan duidelijk worden gemaakt met behulp van een voorbeeld. In figuur 5-10 staat een voorbeeld van een rapportmodel waarin twee hoofdstukken en in totaal drie paragrafen gedefinieerd zijn. De criteria zijn gedefinieerd tussen de <criterium> en </criterium> tags.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE rapportmodel SYSTEM "rapportmodel.dtd">
<rapportmodel>
  <criterium>
    <of>
      <en>
        <FIETS/>
        <PRIJS gd="300"/>
      </en>
      <WIELEN kd="3"/>
    </of>
  </criterium>
  <hoofdstuk>
    <koptekst>Prijsonderzoek</koptekst>
    <paragraaf>
      <koptekst>Racefietsen</koptekst>
    </paragraaf>
  </hoofdstuk>
  <hoofdstuk>
    <criterium>
      <WIELEN is="1"/>
    </criterium>
    <koptekst>Eenwielers</koptekst>
    <paragraaf>
      <koptekst>Voordelen</koptekst>
    </paragraaf>
    <paragraaf>
      <koptekst>Nadelen</koptekst>
    </paragraaf>
  </hoofdstuk>
</rapportmodel>
```

figuur 5-10 XML Rapportmodel eenvoudig

Het eerste criterium in dit rapportmodel heeft betrekking op het gehele rapportmodel. Dit criterium bepaalt of het rapportmodel gebruikt moet worden om het gewenste rapport te genereren. Een criterium bestaat uit een aantal invoervariabelen en grenswaarden samengevoegd met Booleaanse operatoren *en*, *of* en *niet*. De invoervariabelen kunnen afkomstig zijn uit een database of een ander XML bestand. Het criterium voor het rapportmodel in figuur 5-10 gebruikt een drietal invoervariabelen: FIETS, PRIJS en WIELEN. Aan het criterium is voldaan als de invoervariabele FIETS een willekeurige waarde heeft, dus niet leeg of ongedefinieerd is, en de invoervariabele PRIJS een waarde groter dan 300 heeft, of het aantal WIELEN kleiner is dan 3.

In het algemeen kunnen in een criterium naast de invoervariabelen de volgende XML tags voorkomen:

- <of> ... </of>
- <en> ... </en>

Energie Prestatie Advies

- <niet> ... </niet>

Verder kunnen voor de invoervariabelen, te herkennen aan de hoofdletters, de volgende attributen voor de grenswaarde vergelijking gebruikt worden:

- is=" ... ", om de invoervariabele met een grenswaarde te vergelijken;
- kd=" ... ", kleiner dan;
- gd=" ... ", groter dan;
- kgd=" ... ", kleiner dan of gelijk aan;
- ggd=" ... ", groter dan of gelijk aan;
- on=" ... ", ongelijk aan.

Deze vergelijkingsmogelijkheden zouden nog uitgebreid kunnen worden met reguliere expressies en het met elkaar vergelijken van twee invoervariabelen. Reguliere expressies kunnen o.a. gebruikt worden om in tekst te zoeken naar bepaalde patronen. Bij het bedenken van het criterium voor het rapportmodel moet de expert zorgvuldig te werk gaan. Als de expert bijvoorbeeld bepaald heeft dat er drie verschillende rapportmodellen gebruikt moeten gaan worden, moeten de criteria van de drie rapportmodellen elkaar wederzijds uitsluiten. Voor het onderstaande voorbeeld is dit eenvoudig na te gaan.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE rapportmodel SYSTEM "rapportmodel.dtd">
<rapportmodel id="1">
  <criterium>
    <LENGTE kgd="3" />
  </criterium>
</rapportmodel>

<rapportmodel id="2">
  <criterium>
    <en>
      <LENGTE gd="3" />
      <LENGTE kgd="6" />
    </en>
  </criterium>
</rapportmodel>

<rapportmodel id="3">
  <criterium>
    <LENGTE gd="6" />
  </criterium>
</rapportmodel>
```

figuur 5-11 Drie Rapportmodellen (eenvoudig)

Als een criterium niet meer zo eenvoudig is, kan het lastig zijn om te controleren of de criteria een consistent geheel vormen. In het onderstaande voorbeeld overlappen de criteria elkaar zodanig dat het niet duidelijk is welk rapportmodel voor het uiteindelijke rapport gebruikt moet worden. Als LENGTE de waarde 3 heeft, BREEDTE de waarde 6 en DIEPTE de waarde 4, dan voldoen de criteria van alle drie rapportmodellen. Het systeem kan dan bijvoorbeeld kiezen voor het eerste rapportmodel waarvan het criterium voldaan is.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE rapportmodel SYSTEM "rapportmodel.dtd">
<rapportmodel id="1">
  <criterium>
    <LENGTE kgd="3" />
  </criterium>
</rapportmodel>

<rapportmodel id="2">
  <criterium>
    <en>
      <LENGTE ggd="3" />
      <BREEDTE is="6" />
    </en>
  </criterium>
</rapportmodel>

<rapportmodel id="3">
  <criterium>
    <LENGTE is="3" />
    <DIEPTE gd="4" />
  </criterium>
</rapportmodel>
```

figuur 5-12 Drie Rapportmodellen (complex)

Dit probleem kan opgelost worden door het systeem de criteria anders te laten beoordelen. In plaats van de criteria van de rapportmodellen afzonderlijk te controleren en te stoppen zodra een van de rapportmodellen voldoet, kan het systeem alle criteria van alle rapportmodellen controleren en het rapportmodel kiezen waarvan het criterium vervuld is én het meest uitgebreide criterium heeft. Deze methode maakt het voor de expert eenvoudiger om de criteria voor de keuze van het rapportmodel op te stellen, maar dit gaat ten koste van de invloed van de expert op de criteria.

In de hier voorgestelde architectuur zal daarom ook gebruik worden gemaakt van strikte criteria. Deze strikte criteria moeten namelijk ook gebruikt worden voor de keuze van hoofdstukken en paragrafen. Voor deze keuze is het noodzakelijk om strikte criteria te gebruiken. Bij de beslissing uit welke hoofdstukken een rapport bestaat is het niet de keuze tussen verschillende hoofdstukken, vergelijkbaar met de keuze tussen de verschillende rapportmodellen, maar de beslissing of een hoofdstuk in het uiteindelijke rapport geplaatst moet worden van belang. De keuze tussen de mogelijke hoofdstukken is impliciet al gemaakt door de keuze voor een bepaald rapportmodel. Er moet alleen nog de keuze gemaakt worden of een hoofdstuk wel of niet geplaatst moet worden in het uiteindelijke rapport. Deze keuze maakt het gebruik van strikte criteria noodzakelijk. De keuze voor de paragrafen binnen de hoofdstukken zou in principe op dezelfde manier gemaakt kunnen worden als de keuze voor de hoofdstukken, dus met gebruik van strikte criteria.

5.9 Rapportmodel voorbeeld

In de vorige paragraaf is het gebruik van de rapportgenerator beschreven. Deze paragraaf behandelt middels een voorbeeld het genereren van een rapport op basis van een rapportmodel. Op basis van het volledige rapportmodel-DTD is een voorbeeld rapportmodel gemaakt.

```

<?xml version="1.0" encoding="UTF-8"?>
<DOCTYPE rapportmodel SYSTEM "rapportmodel.dtd">
<rapportmodel>
  <criterium>
    <or>
      <PERIODE_ONDERZOEK is="1990-2000"/>
      <PERIODE_ONDERZOEK is="1980-1990"/>
    </or>
  </criterium>
  <hoofdstuk>
    <koptekst>Economische groei en milieueffecten</koptekst>
    <paragraaf>
      <koptekst>Milieu en economie</koptekst>
      <tekstblok>
        Productie en consumptie beïnvloeden de kwaliteit van de natuurlijke omgeving. In de vorm van aantasting van het milieu en uitputting van grondstoffen is die invloed dikwijls negatief. Dit suggereert dat een groei van de economie tot meer aantasting van het milieu leidt en, in samenhang daarmee, tot een afname van de kwaliteit van de natuurlijke omgeving. Het verband tussen economische groei en milieu is echter niet zo strikt. Zo zijn de afgelopen drie decennia, bij een voortgaande groei van productie en consumptie, over een breed front milieuproblemen opgelost of beheersbaar gemaakt.
      </tekstblok>
      <tekst>De waterkwaliteit is sterk <WATERKWALITEIT_VERANDERING/>.</tekst>
      <tekst>De afvalproblematiek lijkt <AFVALPROBLEMATIEK_BEHEERSBAAR/> geworden.</tekst>
      <tekst>De acute luchtverontreiniging door industriële emissies is <LUCHTVERONTREINIGING_BEHEERSBAAR/>.</tekst>
      <tekstblok>
        Met andere woorden: de groei van de economie is voor een deel ontkoppeld van de ongewenste milieueffecten. Daarnaast zijn er echter hardnekkige milieuknelpunten waarvoor de oplossing nog niet binnen bereik is, waaronder de problematiek van de broeikasgassen, verzurende stoffen en aantasting van de leefomgeving door slechte luchtkwaliteit vanwege fijn stof.
      </tekstblok>
    </paragraaf>
    <paragraaf>
      <criterium>
        <PERIODE_ONDERZOEK is="1980-1990"/>
      </criterium>
      <koptekst>Ontwikkelingen in de periode <PERIODE_ONDERZOEK/>
      </koptekst>
      <tekstblok>
        Over de periode tussen 1980 en 1990 is weinig bekend.
      </tekstblok>
    </paragraaf>
    <paragraaf>
      <criterium>
        <PERIODE_ONDERZOEK is="1990-2000"/>
      </criterium>
      <koptekst>Ontwikkelingen in de periode <PERIODE_ONDERZOEK/>
      </koptekst>
      <tekstblok>
        De samenhang tussen economie en milieu is dus een complexe. Alvorens scenario-uitkomsten voor de periode 2003-2010 te presenteren waarin die samenhang verder wordt uitgewerkt, is het, als referentie voor die uitkomsten, interessant om enige relevante ontwikkelingen van de afgelopen periode te presenteren.
      </tekstblok>
      <tekst>Omdat in dit verband vooral structurele ontwikkelingen van belang zijn, is hiervoor de periode <PERIODE_ONDERZOEK/> gekozen.</tekst>
    </paragraaf>
  </hoofdstuk>
  <hoofdstuk>
    <criterium>
      <and>
        <VERVOER/>
        <PERIODE_ONDERZOEK is="1990-2000"/>
      </and>
    </criterium>
    <koptekst>Vervoer</koptekst>
    <tekst>De bedrijfstak transport bestaat uit <ONDERDELEN_TRANSPORT/>.</tekst>
    <tekst>De bedrijfstak transport bestaat niet uit <ONDERDELEN_TRANSPORT_NIET/>.</tekst>
  </hoofdstuk>
</rapportmodel>

```

figuur 5-13 Rapportmodel voorbeeld

De gedefinieerde invoervariabelen en hun waarden staan in de onderstaande tabel.

Invoervariabele	Waarde
PERIODE_ONDERZOEK	1990-2000
WATERKWALITEIT_VERANDERING	verbeterd
AFVALPROBLEMATIEK_BEHEERSBAAR	beheersbaar
VERVOER	vervoer
ONDERDELEN_TRANSPORT	openbaar vervoer
ONDERDELEN_TRANSPORT	goederen vervoer over de weg
ONDERDELEN_TRANSPORT	zeevaart
ONDERDELEN_TRANSPORT	binnenvaart
ONDERDELEN_TRANSPORT	vervoer door de lucht
ONDERDELEN_TRANSPORT	reisbureaus
ONDERDELEN_TRANSPORT_NIET	het eigen vervoer van andere dan transportbedrijven

tabel 5-3 Invoervariabelen

Op basis van dit rapportmodel en de gegeven invoervariabelen kan het rapport van gegenereerd worden.

1 Economische groei en milieueffecten

1.1 Milieu en economie

Productie en consumptie beïnvloeden de kwaliteit van de natuurlijke omgeving. In de vorm van aantasting van het milieu en uitputting van grondstoffen is die invloed dikwijls negatief. Dit suggereert dat een groei van de economie tot meer aantasting van het milieu leidt en, in samenhang daarmee, tot een afname van de kwaliteit van de natuurlijke omgeving. Het verband tussen economische groei en milieu is echter niet zo strikt. Zo zijn de afgelopen drie decennia, bij een voortgaande groei van productie en consumptie, over een breed front milieuproblemen opgelost of beheersbaar gemaakt. De waterkwaliteit is sterk verbeterd en de afvalproblematiek lijkt beheersbaar geworden. Met andere woorden: de groei van de economie is voor een deel ontkoppeld van de ongewenste milieueffecten. Daarnaast zijn er echter hardnekkige milieuknelpunten waarvoor de oplossing nog niet binnen bereik is, waaronder de problematiek van de broeikasgassen, verzurende stoffen en aantasting van de leefomgeving door slechte luchtkwaliteit vanwege fijn stof.

1.2 Ontwikkelingen in de periode 1990-2000

De samenhang tussen economie en milieu is dus een complexe. Alvorens scenario-uitkomsten voor de periode 2003-2010 te presenteren waarin die samenhangen verder worden uitgewerkt, is het, als referentie voor die uitkomsten, interessant om enige relevante ontwikkelingen van de afgelopen periode te presenteren. Omdat in dit verband vooral structurele ontwikkelingen van belang zijn, is hiervoor de periode 1990-2000 gekozen.

2 Vervoer

De bedrijfstak transport bestaat uit openbaar vervoer, goederenvervoer over de weg, zeevaart, binnenvaart, vervoer door de lucht en reisbureaus, maar bestaat niet uit het eigen vervoer van andere dan transportbedrijven.

figuur 5-14 Rapport

Wat opvalt is dat er één paragraaf niet in het uiteindelijk rapport verschenen is, omdat voor die paragraaf niet aan alle criteria voldaan is. Verder zijn er een aantal zinnen met invoervariabelen samengevoegd met het voegwoord *en*. Ook is er een opsomming van een samengestelde invoervariabele geplaatst en is die samengevoegd met een andere opsomming met behulp van het voegwoord *maar*.

Hoofdstuk 6

Ontwerp

In het vorige hoofdstuk is beschreven hoe een EPA geautomatiseerd uitgevoerd kan worden. Dit hoofdstuk richt zich op het ontwerp van het softwaresysteem voor de geautomatiseerde EPA uitvoering. Hiervoor zal een functioneel en technisch ontwerp opgesteld moeten worden. Het volledige ontwerp zal hier niet besproken worden, slechts de opzet, de aandachtspunten en een indicatie van de inhoud wordt gegeven.

6.1 Functioneel Ontwerp

In het functioneel ontwerp wordt in woorden en diagrammen de door het systeem te bieden functionaliteit beschreven. Daartoe wordt allereerst vastgelegd welke actoren een rol spelen bij het gebruik van die functionaliteit. Vervolgens wordt de benodigde functionaliteit per actor vastgelegd. Dit functioneel ontwerp zal opgesteld met behulp van de modelleringstaal UML (Unified Modelling Language). UML wordt gebruikt voor het objectgeoriënteerde modelleren van systemen.

6.1.1 Omgeving

Allereerst zal de gebruikersomgeving van het te ontwerpen systeem vastgelegd moeten worden. Voor het geautomatiseerde EPA systeem zijn dit (gedeeltelijk):

- De ondernemer als representant van de onderneming, ook gebruiker te noemen.
- De EPA adviseur die het advies aan de ondernemer verstrekt, kortweg adviseur.
- De beheerder van het systeem, kortweg beheerder.

Vervolgens zal een opsplitsing van het systeem in functionele deelcomponenten gemaakt moeten worden en zullen de componenten toegelicht moeten worden. Voor het geautomatiseerde EPA systeem zijn dit (gedeeltelijk):

- Gebruikerscomponent;
- Adviseurscomponent;
- Beheerderscomponent.

6.1.2 Functionaliteit

Voor ieder van de beschreven actoren wordt puntsgewijs de door de functionele componenten te bieden functionaliteit beschreven. Dit is gedeeltelijk hieronder weergegeven voor drie genoemde actoren:

Het systeem biedt de gebruiker de volgende functionaliteit:

- Inloggen met persoonlijke inloggegevens.
- Controleren of zijn onderneming in aanmerking komt voor een EPA.

Het systeem biedt de EPA adviseur de volgende functionaliteit:

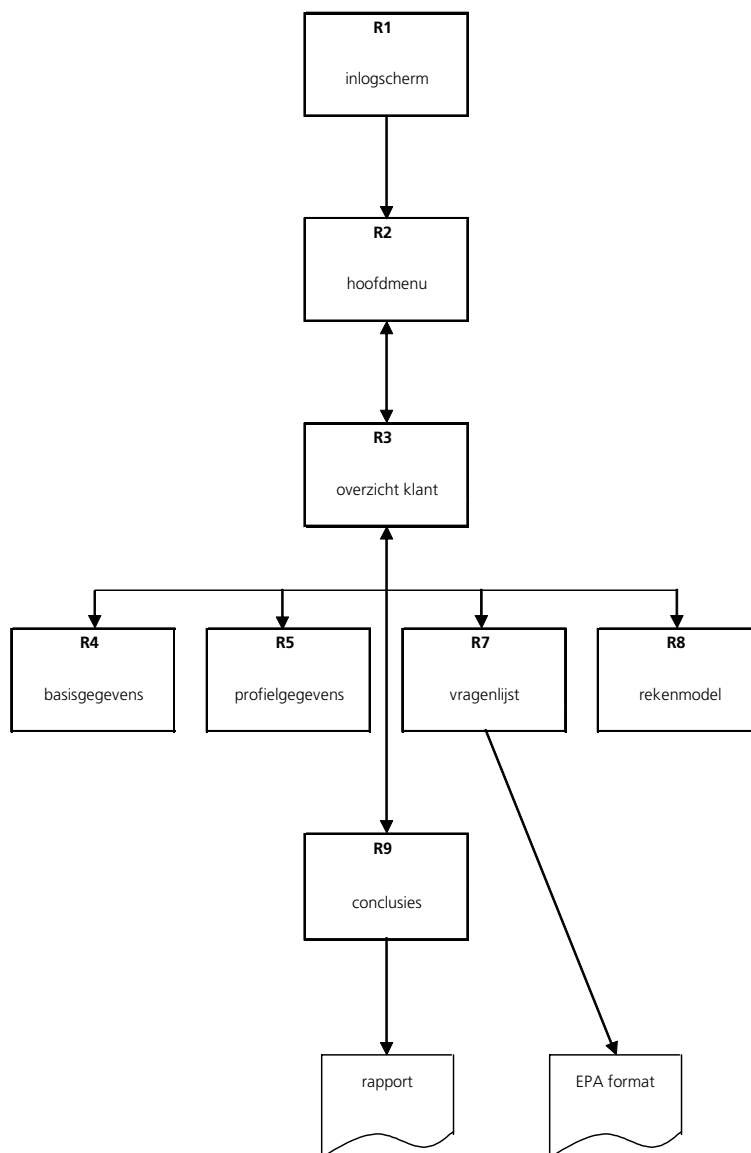
- Inloggen met persoonlijke inloggegevens.
- Overzicht krijgen van gebruikers aan wie hij een EPA advies verstrekt.
- Van die gebruikers kunnen de volgende gegevens beheerd worden (zie vorige hoofdstuk):
 - basisgegevens,
 - profielgegevens,
 - in vragenlijsten ingevulde gegevens,
 - de gegevens in het rekenmodel.
- Een EPA rapport laten genereren op basis van de genoemde gegevens.

Het systeem biedt de beheerder de volgende functionaliteit:

- Beheren van adviseurs:
 - toevoegen,
 - wijzigen,
 - verwijderen.
- Beheren van gebruikers.
- Beheren van de vragen zoals gesteld in de diagnosefase.
- Beheren van de conclusies zoals getrokken uit de diagnosefase.

6.1.3 Stroomschema

Voor webgebaseerde systemen kan de werking van een systeem goed vastgelegd worden met behulp van een stroomschema. Met zo'n stroomschema wordt beschreven welke schermen (webpagina's) en in welke volgorde een actor voor zich krijgt, op basis van de in de vorige paragraaf beschreven functionaliteit. Voor de actor 'adviseur' ziet dit er als volgt uit:



figuur 6-1 Stroomschema Adviseur

Uit dit stroomschema blijkt dat de adviseur eerst moet inloggen alvorens in een hoofdmenu terechtkomt. Van daaruit heeft hij de mogelijkheid om een overzicht te krijgen van zijn klanten (de gebruikers aan wie hij een EPA advies verstrekt)

6.1.4 Schermschetsen

Een preciezere invulling van de in een stroomschema vastgelegde schermen wordt beschreven met behulp van schermschetsen. In deze schermschetsen wordt vastgelegd welke informatie er aangeboden wordt en welke mogelijkheden er geboden worden.

Schermschets R3: 'overzicht klant' en R5: 'profielgegevens' zijn hieronder weergegeven:

R3

klanten overzicht voor [adviseur]

klantnr	naam	adviseur	status
13	Boven & Zn BV	Jansen	voltooid
14	DGS Arbo	Jansen	diagnose
15	BMA Ydo	Jansen	uitvoering
16	Assist BV	Jansen	rapportage

uitloggen
nieuwe klant

figuur 6-2 Schermschets R3

R5

profielgegevens

klantnr 14

bedrijfsnaam DGS

branche Horeca

sector

restaurants	V
> 25.000 kWh	V
> 25.000 m3	V
tussen 1970 en 1980	V

electraverbruik

gasverbruik

bouwjaar pand

opslaan
terug
hoofdmenu

figuur 6-3 Schermschets R5

Deze schermschetsen leggen niet vast hoe een scherm er daadwerkelijk komt uit te zien, er wordt slechts vastgelegd welke informatie in welke vorm gepresenteerd wordt. Voor de kleur van bijvoorbeeld de achtergrond en de knoppen wordt een grafisch ontwerp gemaakt.

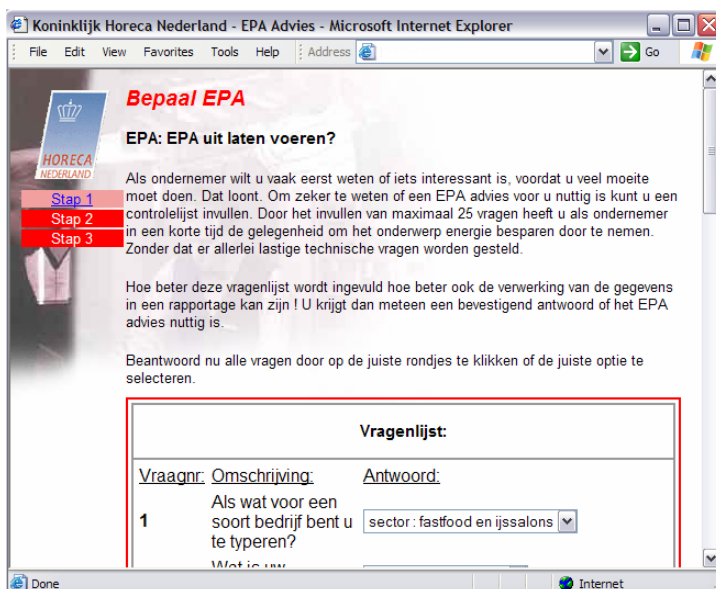
6.2 Grafisch Ontwerp

Het grafisch ontwerp beschrijft in woorden en figuren hoe de schermen zoals vastgelegd in het functioneel ontwerp er daadwerkelijk uit zien. Een voorbeeld van een scherm in het grafisch ontwerp is hieronder weergegeven. Dit is het inlogscherm voor de ondernemer (gebruiker):



figuur 6-4 Inlogscherm Gebruiker

Na het inloggen verschijnt voor de ondernemer het startscherm waarmee hij snel kan bepalen of een EPA nuttig is in zijn situatie.



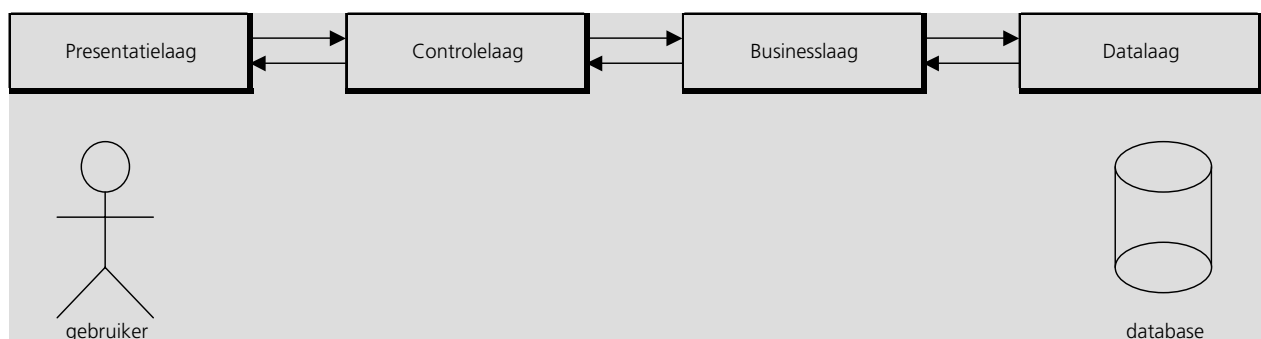
figuur 6-5 Startscherm Gebruiker

6.3 Technisch Ontwerp

Het technisch ontwerp beschrijft de techniek achter de in het functioneel ontwerp gepresenteerde functionaliteit. Er worden in dit technisch ontwerp geen architectuur- of platformspecifieke technieken gebruikt. De gebruikte modelleringstaal, UML, is bovendien programmeertaalafhankelijk. Dit betekent dat het ontwerp in alle (gesimuleerde) objectgeoriënteerde programmeertalen voor internet gerealiseerd kan worden. Er wordt in dit ontwerp een algemene softwarearchitectuur gebruikt waarin onderscheid wordt gemaakt tussen 4 functionele lagen voor de logische scheiding van de verschillende software componenten. Deze lagen luiden als volgt:

- Presentatielaag: Deze laag richt zich op de actoren (gebruiker, beheerder, enz...) Onder de presentatielaag worden de verschillende schermen (ook webpagina's te noemen) verstaan waaruit het systeem is opgebouwd. Deze schermen kunnen dynamisch opgebouwd worden met behulp van een programmeertaal zoals PHP, Java, Perl of Visual Basic. Dit wil zeggen dat de informatie toegespitst kan zijn op de gebruiker, zodat niet iedereen dezelfde informatie, of informatie in een andere vorm, voor zich krijgt.
- Controlelaag: Deze laag bepaalt de acties die ondernomen moeten worden, welke variabelen door de businesslaag (hieronder gelegen laag) verwerkt moeten worden en welke schermen voor de gebruiker naar voren moeten komen.
- Businesslaag: Dit is de laag waarin data bewerkt wordt (indien nodig), berekeningen worden uitgevoerd (rekenmodel) en waarmee data opgeslagen of opgehaald wordt. Deze laag bepaalt grotendeels de functionaliteit van de component.
- Data laag: Dit is de component die zorg draagt voor de opslag van data in het permanente geheugen zodat meerdere gebruikers naast elkaar worden afgehandeld. Deze laag zorgt ervoor dat de database eenduidig benaderd wordt, zodat de verschillende componenten onafhankelijk van elkaar kunnen bestaan.

Het geheel van deze lagen kan in een figuur gepresenteerd worden.



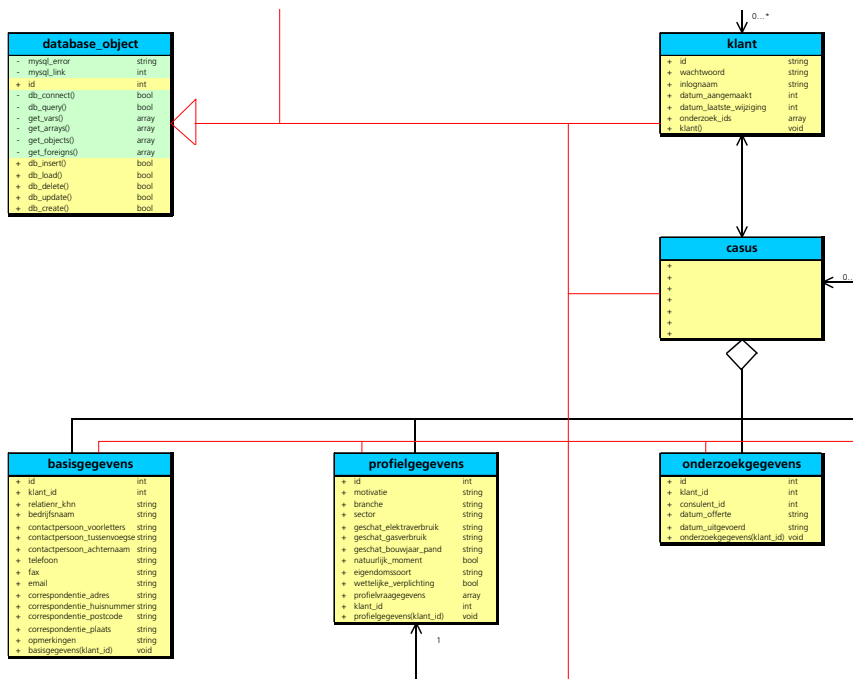
figuur 6-6 Vier lagen architectuur

In bovenstaand figuur is een grafisch weergave van de softwarearchitectuur opgenomen. Er wordt uitgegaan van meerdere componenten bestaande uit een presentatielaag, controlelaag en businesslaag en een uniforme data laag.

6.3.1 Klassendiagram

Een van de meest gebruikte diagrammen in de UML methode is het klassendiagram. In een klassendiagram wordt de structuur aan van een software-intensief systeem vastgelegd. Met de structuur wordt de statische relatie tussen verschillende klassen en interfaces in een systeem bedoeld. Klassen en interfaces zijn termen uit de objectgeoriënteerde programmeertalen. Zij bevatten attributen (variabelen) en methoden. Methoden worden gebruikt voor de communicatie tussen de verschillende klassen.

Voor het geautomatiseerde EPA systeem is een gedeelte van het klassendiagram hieronder weergegeven.

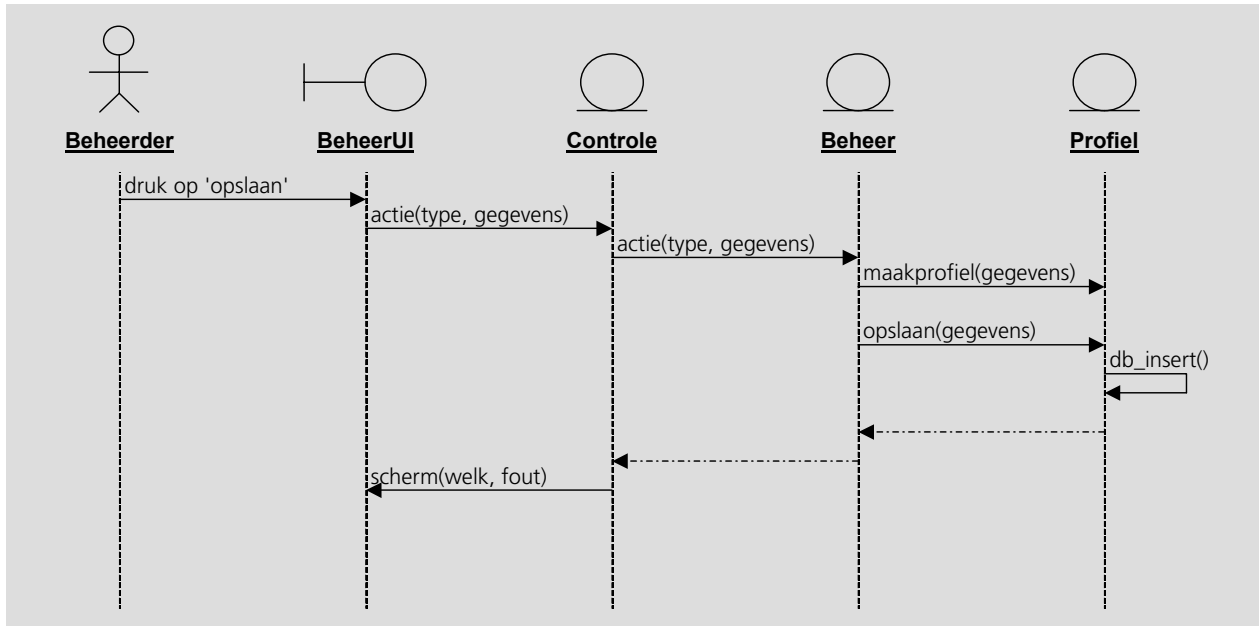


figuur 6-7 Klassendiagram (gedeeltelijk)

Alle klassen van het klassendiagram dienen verder uitgewerkt te worden in het klassenontwerp. Daarin wordt voor iedere klasse beschreven welke attributen en methoden de klasse heeft, waar deze voor gebruikt worden en wat het resultaat van iedere methode is.

6.3.2 Sequentiediagrammen

In Sequentiediagrammen (Sequence Diagrams) wordt de berichtenstructuur en –volgorde tussen de verschillende objecten (concrete instanties van klassen uit het klassendiagram) vastgelegd. Veel van deze sequentiediagrammen zijn gebaseerd op de schermschetsen zoals vastgelegd in het functioneel ontwerp. Voor iedere mogelijke actie van de gebruiker zal een sequentiediagram aangelegd moeten worden. Het sequentiediagram voor de knop 'opslaan' in schermschets R5 is hieronder weergegeven.



figuur 6-8 Sequentiediagram 'R5: opslaan'

6.3.3 Databaseontwerp

Het databaseontwerp, bestaande uit verschillende tabelontwerpen, kan grotendeels afgeleid worden uit het klassendiagram. Alle instanties van klassen die de database_object-interface implementeren zullen opgeslagen worden in de database. Dit betekent dat er voor deze klassen een tabel in de database aangemaakt zal moeten worden. Tevens dienen er tabellen gemaakt te worden voor veel-op-veel (m:n) relaties tussen klassen en voor samengestelde attributen van objecten.

Een voorbeeld van het tabelontwerp voor Profielvraag_antwoord is hieronder weergegeven.

Attribuutnaam	Type	Sleutel	Leeg	Opmerkingen
ID	bigint	primair	nee	Unieke sleutel.
PROFIELVRAAG_ID	bigint	secundair	nee	Verwijzing naar unieke sleutel van profielvraag waartoe dit antwoord behoort.
VOLGORDE	int	nee	nee	Ordering in antwoordvolgorde.
ANTWOORD	text	nee	nee	Het tekstuele antwoord.

figuur 6-9 Tabelontwerp Profielvraag_antwoord

6.4 Realisatie

Als het volledige ontwerp gemaakt, gecontroleerd en goedgekeurd is, kan het gerealiseerd worden. Dit betekent het aanmaken van de tabellen in de database, het programmeren en testen van de beschreven klassen en functionaliteit. De realisatie van het geautomatiseerde EPA systeem zal geschieden op basis van de Microsoft .NET architectuur. Microsoft .NET bevat een grote verzameling ontwikkelgereedschappen voor software. Hierbij kan gedacht worden aan software voor Windows computers, maar ook voor andere randapparatuur en servers. Met Microsoft .NET is het mogelijk verschillende softwareoplossingen aan elkaar te koppelen, mede door de introductie van een gemeenschappelijke compiler. Dat houdt in dat verschillende programmeertalen kunnen worden vertaald naar één gemeenschappelijke programmacode. De rapportgenerator kan dan gemaakt worden in een programmeertaal als Perl en gebruikt worden in een systeem dat verder in Microsofts eigen programmeertaal C#⁹ geschreven is. De programmeertaal Perl is uitsluitend geschikt voor het be- en verwerken van tekst. C# is een objectgeoriënteerde programmeertaal waarmee een goede scheiding tussen de vier logische lagen, zoals aan het begin van deze paragraaf geschetst, gerealiseerd kan worden. Bovendien wordt voor de programmeertaal C# een uitgebreide verzameling standaardcomponenten, o.a. voor integratie van Microsoft Excel bestanden als objecten, aangeboden. Het EPA rekenmodel in Microsoft Excel formaat kan in C# als component geïntegreerd worden.

6.5 Implementatie

Voordat het gerealiseerde systeem in gebruik genomen kan worden, zullen er instructies en handleidingen voor de mensen die ermee gaan werken, geschreven moeten worden. Ook procedures voor het beheer en het gebruik zullen vastgelegd moeten worden.

6.6 Discussie

In dit hoofdstuk is kort de opzet voor en de aandachtspunten van het (software) ontwerp van het geautomatiseerde EPA systeem besproken.

[discussiepunten invoegen](#)

⁹ C#: Engels: "See Sharp"

Hoofdstuk 7

Toepassingsmogelijkheden

In dit hoofdstuk worden verdere toepassingsmogelijkheden van de rapportgenerator en het geautomatiseerde EPA systeem beschreven.

7.1 Uitbreiding naar andere branches

In dit document is bij de beschrijving van het geautomatiseerde EPA systeem en de daarbij te gebruiken rapportgenerator uitgegaan van ondernemingen in de horecabranche. Indien de inzet van deze in de horecabranche een succes blijkt te zijn, ligt een uitbreiding naar andere branches voor de hand. In het ontwerp van het geautomatiseerde EPA systeem is al rekening gehouden met deze mogelijke uitbreiding. Uitbreiding naar andere branches vergt vooral aanpassingen aan de gevraagde informatie, deze zijn grotendeels onafhankelijk van het ontwerp. Ook de rapportgenerator behoeft nauwelijks aanpassingen: Voor het genereren van EPA rapporten voor andere branches kunnen, indien nodig, nieuwe rapportmodellen opgesteld worden.

7.2 Energiecentrum: doe-het-zelf EPA

Het Energiecentrum is een dienstverlenend projectbureau voor kennis- en informatie-overdracht over energie-efficiency en energiebesparing. Brancheorganisaties kunnen een beroep doen op het Energiecentrum voor voorlichting aan hun leden. Bij de overdracht van kennis en informatie maakt het Energiecentrum niet alleen gebruik van de intermediaire rol die de brancheorganisaties naar hun leden vervullen, maar wordt ook direct met deze organisaties samengewerkt. Daarnaast verzamelt en bundelt het Energiecentrum informatie, kennis en capaciteit op het gebied van energie en worden praktische hulpmiddelen ontwikkeld. Daarbij wordt ernaar gestreefd de diensten, activiteiten en hulpmiddelen zoveel mogelijk te laten aansluiten op de belangen en behoeften van de ondernemers.

Een van de mogelijk te ontwikkelen energie-instrumenten is een geautomatiseerde "doe-het-zelf EPA". In de tot dusver besproken uitvoering van een EPA speelde de EPA adviseur een belangrijke rol. In de doe-het-zelf EPA wordt de rol van de EPA adviseur, voor zover mogelijk, vervangen door een software systeem. Dit betekent dat een ondernemer direct zelfstandig een (eenvoudig) EPA rapport kan maken na het invoeren van de noodzakelijke energiegegevens van zijn onderneming.

Hoofdstuk 8

Conclusies en aanbevelingen

In dit afsluitende hoofdstuk worden de conclusies weergegeven zoals die uit deze scriptie voortvloeien. Hierbij wordt een antwoord gegeven op de in hoofdstuk 1 vastgelegde probleemstelling. De onderzoeksvragen zijn in voorgaande hoofdstukken middels een literatuurscriptie en een voorstel voor een geautomatiseerd systeem beantwoord. Vervolgens worden in dit hoofdstuk nog enkele aanbevelingen gedaan met betrekking tot het voorgestelde systeem.

8.1 Conclusies

De probleemstelling die in deze scriptie centraal heeft gestaan, luidt als volgt:

Welke methoden en technieken kunnen er gebruikt worden om een systeem te maken dat zelfstandig kan bepalen of een ondernemer verplicht is, dan wel verstandig er aan doet om een EPA onderzoek te laten uitvoeren en tevens welke methoden en technieken er vervolgens gebruikt kunnen worden om het systeem een EPA rapport te laten genereren aan de hand van de onderzochte gegevens bij de onderneming?

De uitwerking van deze probleemstelling heeft aan de hand van de volgende onderzoeksvragen plaatsgevonden:

- Welke methoden en technieken zijn er om teksten te genereren op basis van kennis uit het kennisdomein van een tekstschrijver?
- Hoe kunnen tekstgeneratietechnieken gebruikt worden in een systeem om rapporten te genereren?
- Hoe kan een rapportgenerator toegepast worden voor het genereren van een EPA rapport?
- Welke informatie en berekeningen zijn er nodig om te komen tot een EPA rapport?
- Hoe kan een rapportgenerator zelfstandig ingezet worden om zonder eigen bijdrage van een expert een niet van handgeschreven rapport te onderscheiden rapport te schrijven?

Conclusies m.b.t. EPA, 3 tekstgeneratietechnieken, rapportgenerator, rapportmodellen, EPA kennis, klantprofiel, EPA geautomatiseerd, EPA rekenmodel, modelleringswijze, ontwerp, uitbreidingsmogelijkheden, rol EPA adviseur

8.2 Aanbevelingen

Mogelijkheden verder onderzoek tekstgeneratietechnieken, verdere ontwikkeling van de rapportgenerator, meer tekstgeneratietechnieken in rapportgenerator, uitbreiding rapportmodel, meer epa rapporten omzetten in rapportmodel, volledige realisatie geautomatiseerd systeem, toepassing in Energieloket indien goedwerkend, inzet in meerdere branches.

Hoofdstuk 9

Referenties

- [1] Ministerie van Volkshuisvesting Ruimtelijke Ordening en Milieubeheer en Ministerie van Economische Zaken, *Circulaire Energie in de milieuvergunning*, Den Haag, oktober 1999.
- [2] Ehud Reiter en Robert Dale, *Building Applied Natural Language Generation Systems*, Cambridge University Press, Cambridge, Engeland, mei 1997.
- [3] Richard I Kittredge en Alain Polguère, *The Generation of Reports from Databases*, University of Montreal, Quebec, Canada, 1994.
- [4] Robert Dale, *Handbook of Natural Language Processing*, Marcel Dekker inc, New York, VS, 2000.
- [5] Karen Kukich, *Design of a Knowledge-Based Report Generator*, University of Pittsburgh, VS.
- [6] Kathleen R McKeown, *The TEXT System for Natural Language Generation: An Overview*, Proceedings of the Twentieth Annual Meeting of the Association for Computational Linguistics, Toronto, Canada, 1982.
- [7] Tim Bray e.a., *Extensible Markup Language (XML) 1.0 (Second Edition)*, <http://www.w3.org/TR/REC-xml>, 2000.
- [8] Ralph Bergmann, *Introduction to Case Based Reasoning*, <http://www.cbr-web.org/CBR-Web/cbrintro/sld001.html>, 1998.
- [9] William Mann e.a., *Text Generation*, American Journal of Computational Linguistics volume 8, april 1982.
- [10] Richard Wheeler, *I KNEW THAT: An Introduction to Case-Based Reasoning*, Edinburgh, 1998.
- [11] James C Shaw, *Clause Aggregation: An Approach to Generating Concise Text*, Columbia University, 2002.
- [12] Henry S Thompson, *Strategy and tactics: A model for language production*, Papers from the 13th Regional Meeting of the Chicago Linguistics Society, 1977.
- [13] Jacques Robin, *Revision-Based Generation of Natural Language Summaries providing Historical Background*, Columbia University, Columbia University, New York, 1995.
- [14] Jacques Robin, *Automatic generation and revision of natural language summaries providing historical background*, Proceedings of the 11th Brazilian Symposium on Artificial Intelligence (SBIA'94), Fortaleza, Brazilië, 1994.
- [15] James Shaw, Shimei Pan en Kathleen McKeown, *An Applied Language Generation System for Intensive Care*, Columbia University, New York, 1996.
- [16] Kathleen McKeown, *MAGIC Seen from the perspective of RAGS*, <http://www.itri.bton.ac.uk/projects/rags/nov-papers/mckeown.ppt>.
- [17] Kathleen McKeown e.a., *Language Generation for Multimedia Healthcare Briefings*, Columbia University, New York, 1997.
- [18] M Dalal e.a., *Negotiation for Automated Generation of Temporal Multimedia Presentations*, Proceedings ACM Multimedia (pag. 55-64), Boston, November 1996.
- [19] Paul H Grice, *Logic and conversation, Syntax and Semantics volume 3: Speech Acts*, Academic Press, New York, 1975.
- [20] Michael Halliday en R Hasan, *Cohesion in English*, Londen, 1976.

- [21] Rebecca Passonneau e.a., *Generating summaries of workflow diagrams*, Proceedings of the International Conference on Natural Language Processing and Industrial Applications, New Brunswick, Canada, 1996.
- [22] Agnar Aamodt en Enric Plaza, *Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches*, *Artificial Intelligence Communications* 7, 1994.
- [23] Hans van Mourik en Ivo Zandhuis, *XML Toepassingen, SURFworks Rapportage*, Leiden, 2002.
- [24] Ian Watson, *Survey of CBR Application Areas*, www.ai-cbr.org, 1999.
- [25] Taalunie, *Woordenlijst Nederlandse taal (Groene Boekje)*, SDU, 's Gravenhage, 1998.

Bijlage I BMA

De BMA Groep is een professionele dienstverlenende organisatie die zich bezighoudt met advisering, kennisoverdracht en projectuitvoering binnen profit, non-profit en overheidssectoren op talrijke gebieden. De BMA Groep is opgericht in 1990 en van oorsprong een organisatieadviesbureau.

De BMA Groep streeft complete dienstverlening na. Binnen de BMA Groep staat maximale persoonlijke aandacht voor de relaties voorop. Vanuit de BMA Groep wordt intensieve samenwerking tussen de diverse disciplines gestimuleerd en wordt de synergie voortdurend bewaakt. Ook worden tal van taken en diensten centraal gecoördineerd en/of uitgevoerd.

De huidige organisatiestructuur van de BMA Groep past geheel binnen de basisprincipes die oprichter Cor Baars in 1990 voor ogen had. Deze principes staan vandaag de dag nog altijd onverkort overeind; alles draait om openheid, eerlijkheid en gelijkwaardigheid. Deze principes worden ook door de afzonderlijke ondernemingen op de relaties overgedragen. In de benadering van de cliënten blijkt het voordeel dat deel wordt uitgemaakt van een grote, breed opererende onderneming: de BMA Groep.

BMA Advies ondersteunt groepen van MKB bedrijven, brancheorganisaties, intermediairs en overheden, met betrekking tot praktische vraagstukken in het kader van o.a.: wet- en regelgeving, arbo- en veiligheid en maatschappelijk verantwoord ondernemen.

BMA Advies heeft naam gemaakt als deskundig en innovatief adviesbureau, met een voorkeur voor praktische oplossingen, flexibiliteit en op zijn tijd een iets wat onorthodoxe aanpak. Kenmerkend is de combinatie van een relatie- en resultaatgerichte werkwijze.

BMA Multimedia integreert creativiteit, vormgeving, techniek en functionaliteit in complete maatwerkproducten. BMA Multimedia verleent als werkmaatschappij van de BMA Groep haar diensten op het gebied van multimediacommunicatie, voornamelijk aan MKB-brancheorganisaties en (semi-)overheid.

BMA Multimedia is gespecialiseerd in het verzorgen van totaalproducties, van conceptontwikkeling tot en met ontwerp, productie, implementatie, onderhoud en nazorg.