TECHNISCHE HOGESCHOOL DELFT
Afdeling der Elektrotechniek
Vakgroep Automatische Verkeerssystemen
Mekelweg 4
Postbus 5031
2600 GA Delft

a f s t u d e e r o p d r a c h t

Te verrichten door: J.C. van der Lippe

De opdracht zal worden uitgevoerd op het dr. Neher Laboratorium.

Mentoren: D. Sparreboom
N.C.J.M. de Beer

Omschrijving van de opdracht:

Ontwerp en maak een programma dat trefwoorden uit een Viditel-pagina afleidt. Geef een gemotiveerd oordeel over de geschiktheid van dit programma als hulpmiddel bij het toekennen van trefwoorden aan Viditel-pagina's.
Als blijkt dat het programma daarbij goede hulp kan bieden, ontwerp dan een dialoog via welke zulke trefwoorden als suggestie bij het toekennen van trefwoorden worden aangeboden. De programmatuur dient zo gemaakt te worden dat inpassing in bestaande Viditel-programmatuur eenvoudig mogelijk is.

Begin van de opdracht:
2 september 1985.

Delft, september 1985.

b.g.

prof. ir. J.L. de Kroes

aantal pagina's : 51
titel : computer-ondersteund indexeren van Viditelpagina's
schrijver : J.C. v.d. Lippe
samenvatting :

Viditel beschikt over een grote databank die informatie bevat over uiteenlopende onderwerpen. De informatie bevindt zich op zogenaamde pagina's. VAKIR, een nieuwe zoekmethode voor Viditel, stelt gebruikers in staat om hun informatiebehoefte kenbaar te maken via zelf gekozen trefwoorden.

Voor het functioneren van VAKIR is het noodzakelijk dat de leverancier van de informatie trefwoorden aan zijn pagina's toekent. Om de informatieleverancier bij dit zogenaamde 'indexeren' ondersteuning te bieden is onderzoek verricht naar de mogelijkheden om uit de tekst van een pagina trefwoorden af te leiden. Dit onderzoek heeft geresulteerd in de implementatie van een nieuwe indexeerdialoog, via welke de informatieleverancier suggesties krijgt aangeboden voor aan zijn pagina's toe te kennen trefwoorden.

VOORWOORD

Dit rapport geeft het verslag van het afstudeerwerk dat door J.C. van der Lippe op het Dr. Neher Laboratorium van de nederlandse PTT is verricht. De schrijver wil iedereen die, op welke manier dan ook, een bijdrage heeft geleverd aan het afstudeerwerk hartelijk bedanken.

SUMMARY

Viditel possesses a large database which contains information about numerous subjects. The information is stored on pages. Finding specific information in the database is often a problem owing to the limitations of the existing search methods. Experiments showed that Viditel Advanced Keyword Information Retrieval system (VAKIR) increases the accessibility of the information. VAKIR allows the user to select information by entering freely chosen keywords.

In order for VAKIR to function it is necessary that information providers assign keywords to their pages. This process of assigning keywords is called 'indexing'. In order to give an information provider support with this, research has taken place into the possibilities of deriving keywords from pages. This research has resulted in the implementation of a new indexing dialogue; through which suggestions are given for keywords that could be assigned to a page.

SAMENVATTING

Viditel beschikt over een grote databank die informatie bevat over uiteenlopende onderwerpen. De informatie bevindt zich op zogenaamde pagina's. Een vaak gehoord punt van kritiek op Viditel richt zich op de beperkte mogelijkheden om de gewenste informatie te kunnen bereiken. Experimenten hebben aangetoond dat Viditel Advanced Keyword Information Retrieval system (VAKIR), een nieuwe zoekmethode voor Viditel, de toegankelijkheid van de informatie verbetert. VAKIR stelt gebruikers in staat om hun informatiebehoefte kenbaar te maken via zelf gekozen trefwoorden.

Voor het functioneren van VAKIR is het noodzakelijk dat de leverancier van de informatie trefwoorden aan zijn pagina's toekent. Dit toekennen van trefwoorden aan pagina's wordt indexeren genoemd. Om een informatieleverancier daarbij ondersteuning te kunnen bieden is onderzoek verricht naar de mogelijkheden om uit de tekst van een pagina trefwoorden af te leiden. Dit onderzoek heeft geresulteerd in de implementatie van een nieuwe indexeerdialoog, via welke een informatieleverancier suggesties krijgt aangeboden voor aan zijn pagina's toe te kennen trefwoorden.

LIJST VAN GEBRUIKTE AFKORTINGEN

VAKIR	: Viditel Advanced Keyword Information Retrieval system.
DNL	: Dr. Neher Laboratorium van de Nederlandse PTT.
STL	: Standaard-Trefwoordenlijst.
v	: Volledigheid.
j	: Juistheid.

INHOUDSOPGAVE

BIJLAGE

1. INLEIDING

Viditel beschikt over een grote databank die informatie bevat over uiteenlopende onderwerpen. De informatie bevindt zich op zogenaamde pagina's. Gebruikers willen uit het grote aanbod van pagina's op een snelle en eenvoudige wijze hun keuze kunnen maken. Daartoe komt in de toekomst Viditel Advanced Keyword Information Retrieval system (VAKIR) beschikbaar. VAKIR biedt gebruikers de mogelijkheid om hun informatiebehoefte kenbaar te maken via zelf gekozen trefwoorden.

Voor het functioneren van VAKIR is het noodzakelijk dat de leverancier van de informatie trefwoorden aan zijn pagina's toekent. Dit toekennen van trefwoorden aan pagina's wordt indexeren genoemd. Om een informatieleverancier daarbij ondersteuning te kunnen bieden is onderzoek verricht naar de mogelijkheden om uit de tekst van een pagina trefwoorden af te leiden.

Allereerst worden in hoofdstuk 2 de bestaande zoekmethoden in Viditel behandeld. Ook VAKIR komt aan de orde. Hoofdstuk 3 beschrijft een onderzoek naar de mogelijkheden om trefwoorden uit de tekst van een pagina af te leiden. Hoofdstuk 4 behandelt een nieuwe indexeerdialoog via welke de informatieleverancier suggesties krijgt aangeboden voor aan zijn pagina's toe te kennen trefwoorden. Tenslotte komen in hoofdstuk 5 aanbevelingen ten behoeve van de nieuwe indexeerdialoog aan de orde.

2. ZOEKMETHODEN IN VIDITEL

2.1 Wat is Viditel?

In de jaren zeventig is door medewerkers van het British Post Office onderzoek verricht naar de wijze waarop de computer kon worden gebruikt voor een interactieve informatie-opvraagdienst. De te ontwikkelen dienst moest voor iedereen toegankelijk worden en eenvoudig te bedienen zijn. Een uitgangspunt daarbij was dat in vele gezinnen een televisietoestel en een telefoonaansluiting aanwezig zijn. Door een aangepast televisietoestel dienst te laten doen als 'terminal' en het telefoonnet als communicatienetwerk tussen terminal en databank, zou er een groot aantal potentiële gebruikers onstaan. Tenslotte moest de databank zeer groot zijn om aan de informatiebehoefte van vele verschillende gebruikers te kunnen voldoen.

De onderzoekingen hebben geleid tot het onstaan van 'Prestel'. Prestel kan worden beschouwd als de oervader van de zogenaamde Videotexdiensten. Kenmerkend voor Videotexdiensten is dat tekst en grafische informatie wordt getoond op een beeldscherm. Het nederlandse Viditel is voortgekomen uit Prestel. Viditel is in de loop der tijd uitgebreid met een berichtendienst, een tarievenverrekening en de mogelijkheid tot koppeling met andere computers.

2.2 Bestaande zoekmethoden in Viditel

Voor het opsporen van informatie in de databank van Viditel is een vijftal zoekmethoden beschikbaar. Deze zijn: onderwerpenlijst, systematisch zoeken, trefwoordencatalogus, lijst van informatieleveranciers en de lokale zoekmethode. Van elk volgt hieronder een korte beschrijving.

1. Alfabetische lijst van onderwerpen (onderwerpenlijst)

   Deze werkt als volgt:

   a. De gebruiker neemt één trefwoord in gedachte dat van toepassing is op de vraag waarop hij het antwoord in Viditel wil vinden.

   b. Hij voert de letters van dit trefwoord één voor één in. Dit gebeurt door telkens het getal in te toetsen dat op een menupagina voor de betreffende letter staat vermeld. Zie figuur 1 . Na het intoetsen van één of meer letters verschijnt een lijstje van genummerde trefwoorden, dat al dan niet het gewenste trefwoord bevat.

   c. Na het kiezen van een trefwoord uit het lijstje verschijnt een lijst van informatieleveranciers, die elk informatie leveren over het gekozen onderwerp.

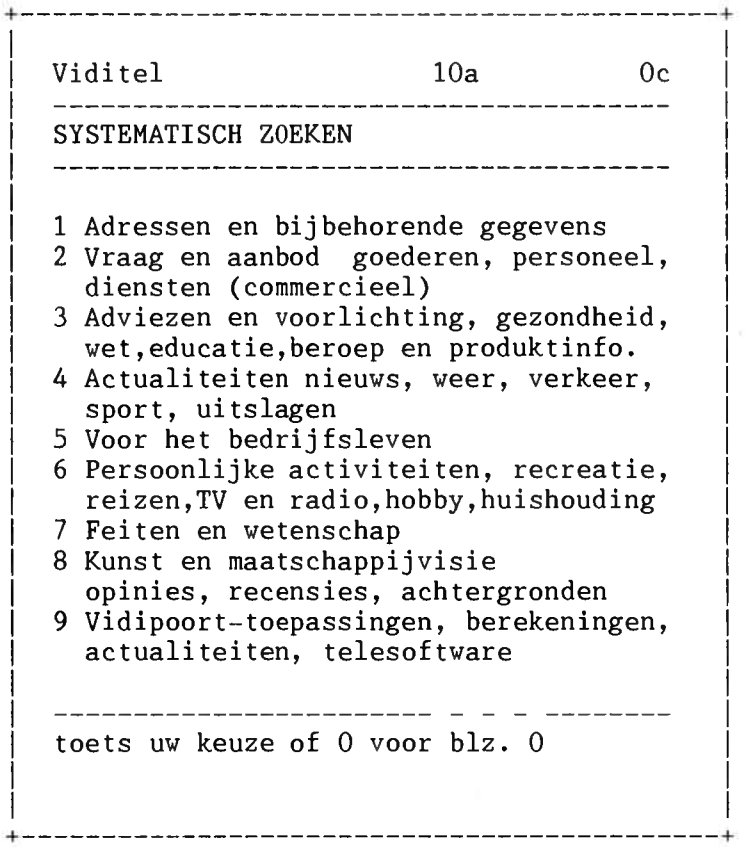
d. Keuze uit het lijstje van informatieleveranciers brengt de gebruiker bij de pagina's van de betreffende informatieleverancier.

```
+-----------------------------------------+
|                                         |
|  Viditel                199685a     0c  |
|  ---------------------------------------|
|  TREFWOORDEN                Sa t/m Sz   |
|  ---------------------------------------|
|                                         |
|                                         |
|                                         |
|                                         |
|  11 Sa      -- Sj       -- Ss           |
|  -- Sb      22 Sk       32 St           |
|  13 Sc      23 Sl       33 Su           |
|  -- Sd      -- Sm       -- Sv           |
|  15 Se      25 Sn       -- Sw           |
|  -- Sf      26 So       -- Sx           |
|  -- Sg      27 Sp       37 Sy           |
|  18 Sh      28 Sq       -- Sz           |
|  19 Si      -- Sr                       |
|                                         |
|                                         |
|                                         |
|                                         |
|  ---------------------- - - - --------  |
|  toets uw keuze of 0 voor index a t/m z |
|                                         |
|                                         |
+-----------------------------------------+
```

Figuur 1 : De onderwerpenlijst

2. Systematisch zoeken

Bij deze zoekmethode maakt de gebruiker steeds een keuze uit een menupagina met informatiecategorieën. Zie figuur 2 . Iedere keuze leidt naar een nieuwe menupagina met meer specifieke informatiecategorieën. Na een aantal keuzes verschijnt een lijst met namen van informatieleveranciers die elk informatie leveren over het gekozen onderwerp. Tenslotte selecteert de gebruiker de pagina's van één van de informatieleveranciers.



Figuur 2 : Systematisch zoeken

3. Trefwoordencatalogus

De gebruiker bepaalt het onderwerp waarover hij wat wil weten en zoekt dat op in een gedrukte trefwoordencatalogus. Na het intoetsen van het paginanummer dat achter het onderwerp staat vermeld, verschijnt een lijst van informatieleveranciers. Vervolgens selecteert de gebruiker de pagina's van één van de informatieleveranciers uit het lijstje.

4. Lijst van informatieleveranciers

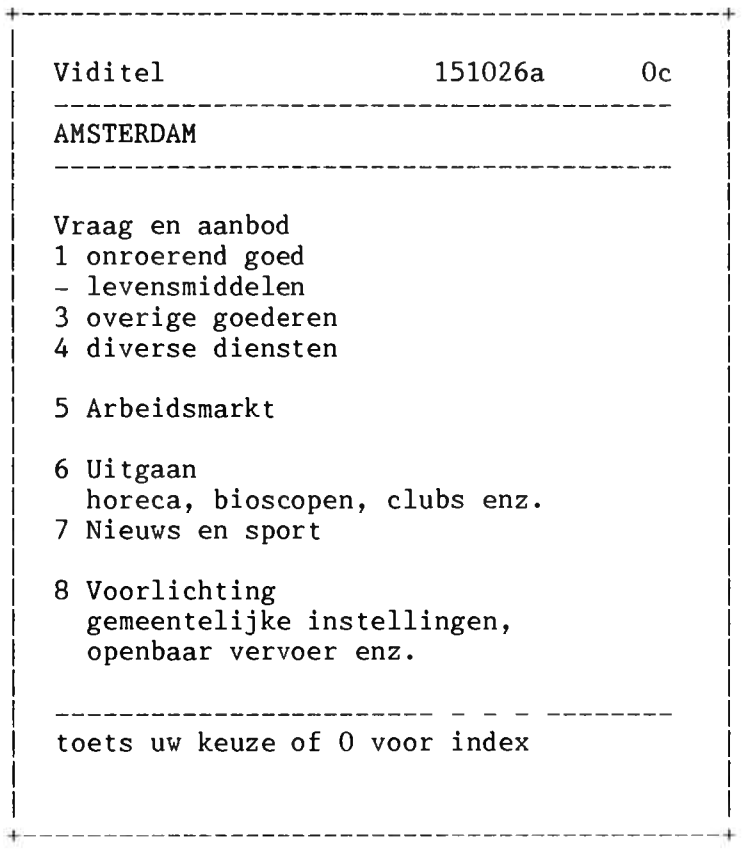
De gebruiker kiest de naam van een bepaalde informatieleverancier en voert deze in zoals bij de onderwerpenlijst. Vervolgens wordt een menupagina met namen van informatieleveranciers getoond. Zie figuur 3 . Door het intoetsen van het paginanummer dat achter elke naam staat vermeld, belandt de gebruiker bij de pagina van de betreffende informatieleverancier.

```
+-----------------------------------------+
|                                         |
|  Viditel              198114a      0c   |
|  ---------------------------------      |
|  INFORMATIELEVERANCIER      HOOFD-      |
|  D                          BLADZIJDE   |
|  ---------------------------------      |
|  DAF Ned.Bedryfswagen BV... 212937      |
|  DAILY SERVICE VIDEOTEX.... 252         |
|  DAN-AIR.................. 256119      |
|  DATASKIL................. 350 paraplu  |
|  Datastream International.. 350574      |
|  DATA VIEW NEDERLAND B.V... 555033      |
|  DATEX SOFTWARE BV........ 44481        |
|  David Computer Systems.... 272094      |
|  Dcw...................... 315 paraplu  |
|  DENKSPORT/DATASKIL....... 35010        |
|  DFDS Fred.Olsen-Bergen.... 2561362     |
|  DFDS Tor Line............ 256136      |
|  Dennis Music............. 2048215     |
|                                         |
|  rechtstreeks kiezen: * nr.hfd.blz. #   |
|  ----------------------- - - - -------  |
|  toets # voor vervolg                   |
|  toets 0 voor alfabetische lijst IL's   |
|                                         |
|                                         |
+-----------------------------------------+
```

Figuur 3 : Lijst van informatieleveranciers

5. Lokale zoekmethode

De lokale zoekmethode wordt gebruikt voor het toegankelijk maken van informatie die is gericht op gebruikers in een bepaalde gemeente en haar directe omgeving. De gebruiker kiest de naam van een gemeente en voert deze in zoals bij de onderwerpenlijst. Als de gekozen gemeente aanwezig is verschijnt een menupagina met per gemeente acht vaste rubrieken. Zie figuur 4 .



Figuur 4 : De lokale zoekmethode

2.3 Ervaringen met de bestaande zoekmethoden

Een vaak gehoord punt van kritiek op Viditel richt zich op de beperkte mogelijkheden om de gewenste informatie te bereiken. In een op het Dr. Neher Laboratorium van de Nederlandse PTT (DNL) gehouden proef [1] is deze kritiek onderzocht. In de proef moesten proefpersonen antwoorden op vragen opsporen met behulp van de volgende zoekmethoden:

- onderwerpenlijst
- systematisch zoeken
- trefwoordencatalogus

- lijst van informatieleveranciers

Op elke vraag was een antwoord in Viditel aanwezig.

Tabel 1 toont de resultaten van de proef. De tijd tussen haakjes is de tijd nodig voor het bestuderen van de trefwoordencatalogus.

	percentage correct beantwoorde vragen	zoektijd (s)
onderwerpenlijst	56.6	370
systematisch zoeken	61.5	318
trefwoordencatalogus	69.0	(50)+256
lijst van informatieleveranciers	36.7	227

Tabel 1 : Resultaten per zoekmethode

Uit tabel 1 volgt dat in een groot aantal gevallen het antwoord op een vraag, alhoewel in Viditel aanwezig, niet met de genoemde zoekmethoden wordt gevonden. Hieruit volgt dat de informatie onvoldoende toegankelijk is via de bestaande zoekmethoden. De bovengenoemde kritiek wordt daarmee bevestigd. Verder kan worden opgemerkt dat zoeken met behulp van de trefwoordencatalogus nog in de meeste gevallen tot het vinden van het juiste antwoord leidt.

De voornaamste tekortkomingen van de bestaande zoekmethoden zijn:

- In de onderwerpenlijst zijn te weinig trefwoorden opgenomen. Het gevolg hiervan is dat door gebruikers gekozen trefwoorden vaak niet tot de gewenste informatie leiden.

- Er zijn verschillen in representaties van informatie door proefpersonen en het systeem. Bij het beantwoorden van vragen zoekt de gebruiker naar de informatiecategorie die volgens hem het meest van toepassing is op de te beantwoorden vraag. Dit is echter niet in alle gevallen de categorie die ook daadwerkelijk tot het vinden van het antwoord leidt.

- Er is te weinig terugkoppeling naar de gebruiker omtrent de juistheid van de gevolgde zoekweg. Vaak blijkt pas in een vrij laat stadium dat de gevolgde zoekweg foutief is en moet weer van voren af aan worden begonnen.

- Onbekendheid bij de gebruiker met de informatie die de verschillende informatieleveranciers aanbieden leidt regelmatig tot de keuze van een verkeerde informatieleverancier.

- Er is een te grote belasting op het geheugen van de gebruiker. Omdat het systeem de door een gebruiker gevolgde zoekweg niet vastlegt moet de gebruiker deze zelf onthouden. Vanwege het grote aantal te onthouden handelingen leidt dit al snel tot overbelasting van het geheugen.

2.4 VAKIR: een nieuwe zoekmethode voor Viditel

Eind 1979 startte op het DNL een studie [2] naar een zoekmethode om de informatie in Viditel beter toegankelijk te maken. Uit deze studie kwam naar voren dat de voornaamste oorzaak voor het slecht toegankelijk zijn van de informatie in Viditel ligt bij de bestaande zoekmethoden. Marktonderzoek wees uit dat op dat moment geen geschikte zoekmethode beschikbaar was. Tesamen met het feit dat gunstige ervaringen waren opgedaan met het selecteren van informatie via trefwoorden, werd geadviseerd om een experimentele trefwoordenzoekmethode te ontwikkelen. Dit advies heeft geleid tot het onstaan van Viditel Advanced Keyword Information Retrieval system (VAKIR). VAKIR stelt gebruikers in staat om hun informatiebehoefte kenbaar te maken door middel van zelf gekozen trefwoorden. VAKIR is geïmplementeerd in Viditest, het test- en ontwikkelsysteem voor Viditelsoftware van het DNL.

VAKIR bestaat functioneel uit drie delen; een deel voor de gebruiker, een deel voor de informatieleverancier en een deel voor de systeembeheerder. Het deel voor de gebruiker vormt de hoofdfunctie van VAKIR en wordt in deze paragraaf behandeld.

Een gebruiker communiceert met VAKIR via een dialoog. Deze dialoog bestaat uit drie fasen.

In de eerste fase, de "definitie-fase", maakt de gebruiker zijn informatiebehoefte aan het systeem kenbaar via het intoetsen van zelf gekozen trefwoorden. Onder een trefwoord wordt verstaan ieder woord of combinatie van woorden waarmee een gebruiker zijn informatiebehoefte omschrijft. VAKIR tracht de ingetoetste trefwoorden te herkennen. Daartoe beschikt het over een lijst van ruim 7500 'hoofdtrefwoorden'. Deze zogenaamde 'Standaard-Trefwoordenlijst' (STL) is tot stand gekomen na overleg met medewekers van de 'Koninklijke Bibliotheek' en het 'Nederlands Bibliotheek en Lectuur Centrum'. Hoofdtrefwoorden zijn trefwoorden met een algemene betekenis. Aan de STL zijn toegevoegd de namen van alle in Viditest aanwezige informatieleveranciers.

Tussen twee trefwoorden in de STL kan een relatie bestaan. De volgende relaties zijn mogelijk:

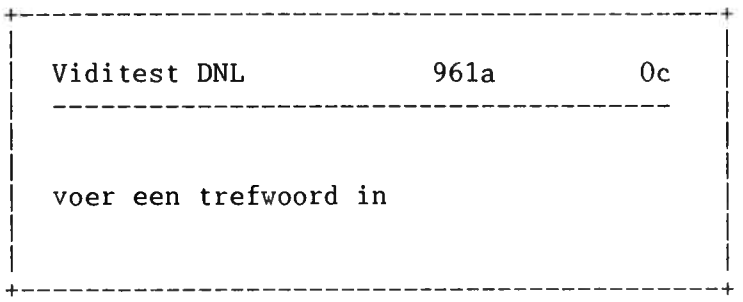
- synoniem relatie: twee trefwoorden hebben dezelfde betekenis.
  Bijvoorbeeld: luchthavens - vliegvelden.

- hiërarchische relatie: het ene trefwoord heeft een beperktere betekenis dan het andere.
  Bijvoorbeeld: volksmuziek - muziek.

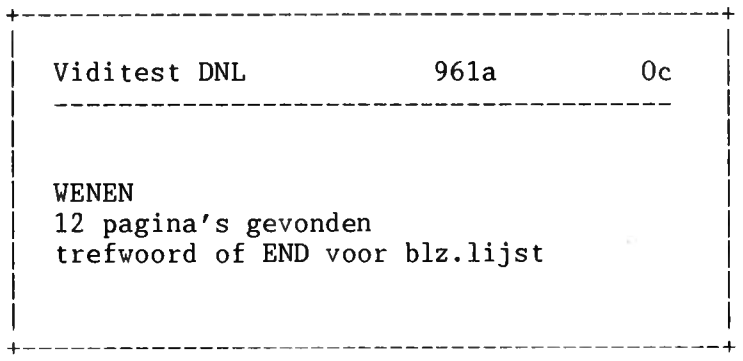
- associatieve relatie: beide trefwoorden omschrijven facetten van een bepaald onderwerp.
  Bijvoorbeeld: fysiotherapie - gezondheidszorg.

Verder kent VAKIR zogenaamde 'subtrefwoorden'. Subtrefwoorden zijn trefwoorden met een specifieke betekenis die is gerelateerd aan één bepaalde informatieleverancier. Iedere informatieleverancier beheert zijn eigen verzameling subtrefwoorden. Gebruikers die de naam van een informatieleverancier intoetsen maken daarmee de verzameling subtrefwoorden van die informatieleverancier toegankelijk. Dat betekent dat VAKIR dan ook die subtrefwoorden herkent.

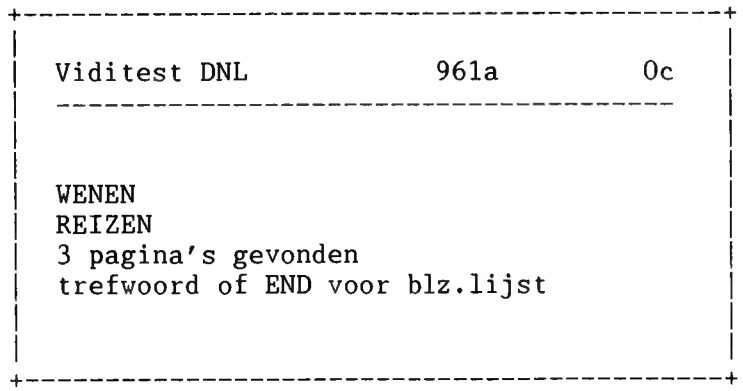
Als VAKIR het ingetoetste trefwoord herkent dan selecteert het de pagina's die betrekking hebben op alle tot dan toe ingetoetste trefwoorden. Zie figuur 5a, 5b en 5c. Tevens meldt het hoeveel pagina's zijn geselecteerd, zodat een goed inzicht wordt verkregen in de voortgang van het zoekproces. Ieder nieuw ingevoerd trefwoord omschrijft de informatiebehoefte van de gebruiker weer nauwkeuriger en reduceert daarmee het aantal geselecteerde pagina's.



Figuur 5a



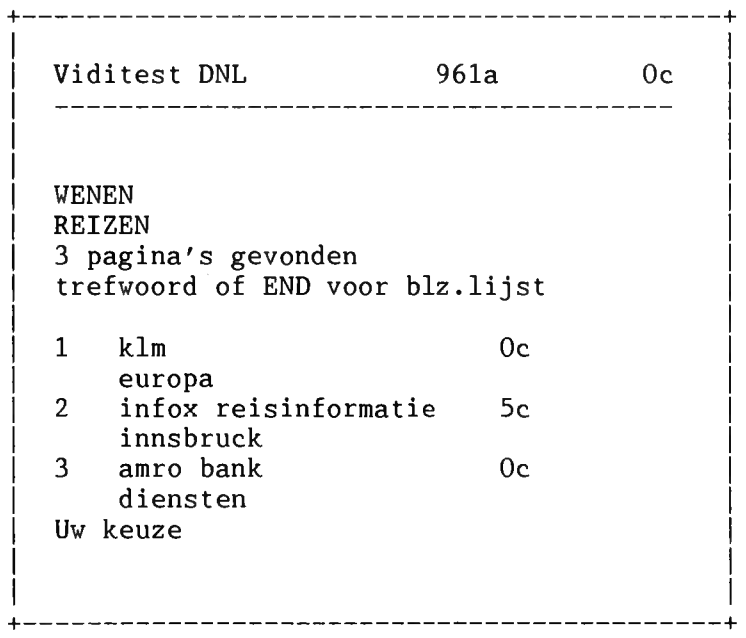
Figuur 5b



Figuur 5c

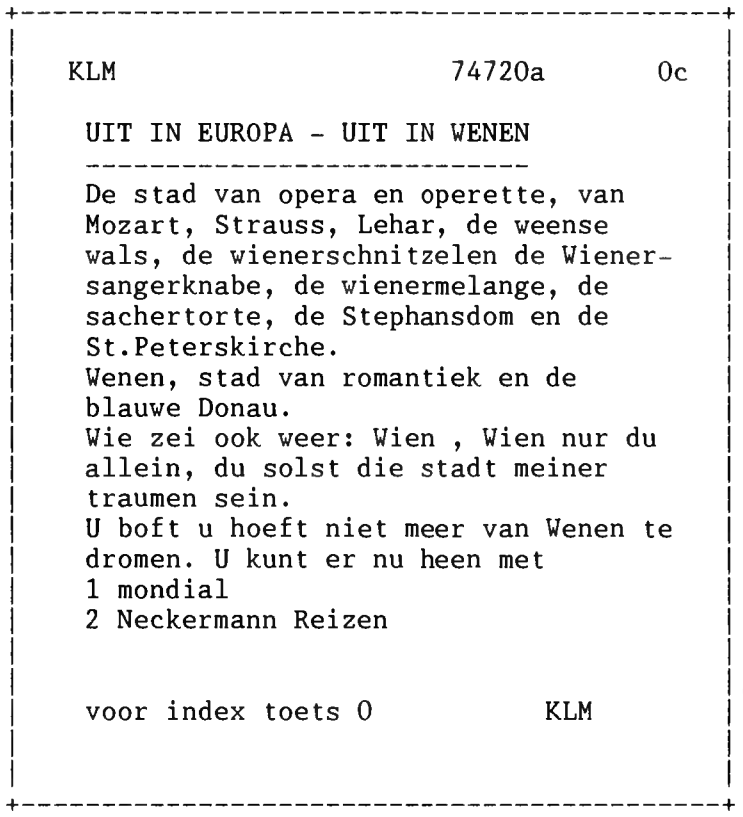
Figuur 5a,b,c : De definitie-fase

De tweede fase, de "selectie-fase", start als de gebruiker vindt dat het aantal geselecteerde pagina's voldoende is gereduceerd. In de tweede fase stelt VAKIR een menupagina samen met daarin een overzicht van de geselecteerde pagina's. Zie figuur 6 . Het overzicht bevat een genummerde lijst van pagina's met voor iedere pagina de naam van de betreffende informatieleverancier, een descriptor die een nadere omschrijving van de inhoud van de pagina geeft en de prijs van de pagina. De descriptor is een aan de pagina gerelateerd hoofd- of subtrefwoord dat nog niet is ingetoetst tijdens fase één. Met behulp van deze informatie kan de gebruiker zijn keuze maken uit het overzicht. Eventueel kan worden teruggekeerd naar fase één van de dialoog.



Figuur 6 : De selectie-fase

In de derde fase, de "raadpleeg-fase", wordt de gekozen pagina aan de gebruiker getoond. Zie figuur 7 . Eventueel kan worden teruggekeerd naar fase twee van de dialoog.



Figuur 7 : De raadpleeg-fase

VAKIR biedt de gebruiker uitgebreide hulpfaciliteiten. Deze hulpfaciliteiten, die worden aangeboden tijdens de dialoog, kunnen worden opgedeeld in drie categorieën.

1. Morfologische hulp. Deze heeft betrekking op de spelwijze van de trefwoorden.

2. Semantische hulp. Deze heeft betrekking op de betekenis van de trefwoorden.

3. Procedurele hulp. Deze heeft betrekking op de bedieningsmogelijkheden van VAKIR.

Elk van de drie hulpvormen kan worden opgedeeld in twee subcategorieën, namelijk impliciete hulp en expliciete hulp. Impliciete hulp wordt aangeboden zonder verder verzoek van gebruiker hierom. Expliciete hulp wordt aangeboden na dat de gebruiker daar, door middel van een commando,

om heeft verzocht.

De volgende zes vormen van hulp zijn mogelijk:

1. impliciete morfologische hulp

   Deze hulp biedt de gebruiker de mogelijkheid om verkorte spelwijzen van trefwoorden toe te passen, staat alternatieve spelwijzen toe en corrigeert in beperkte mate gemaakte spelfouten.

2. expliciete morfologische hulp

   Als een ingetoetst trefwoord niet wordt herkend dan is soms een trefwoord beschikbaar dat qua spelwijze lijkt op het ingetoetste.

3. impliciete semantische hulp

   Als blijkt dat er geen pagina's zijn die betrekking hebben op het ingetoetste trefwoord dan wordt dit trefwoord, indien mogelijk, vervangen door een synoniem ervan dat wel gerelateerde pagina's kent.

4. expliciete semantische hulp

   Op verzoek van de gebruiker levert het systeem een lijst van trefwoorden die in betekenisrelatie staan met het ingetoetste.

5. impliciete procedurele hulp

   Deze vorm van hulp bestaat uit het genereren van diverse boodschappen.

6. expliciete procedurele hulp

   Op elk tijdstip is een overzicht beschikbaar van de in die fase van de dialoog relevante functietoetsen en hun betekenis.

2.5 Ervaringen met VAKIR

Om te onderzoeken of VAKIR de toegankelijkheid van de informatie verbetert, zijn op het DNL laboratoriumproeven [3] gehouden. In deze proeven werd VAKIR vergeleken met één van de bestaande zoekmethoden, namelijk de onderwerpenlijst. Deze laatste methode werd ter vergelijking gekozen om twee redenen:

1. de onderwerpenlijst is een andere zoekmethode waarin gebruik wordt gemaakt van trefwoorden.

2. in beide zoekmethoden wordt geen gebruik gemaakt van externe bronnen, zoals bijvoorbeeld de gedrukte trefwoordencatalogus.

In de proef moesten proefpersonen antwoorden op vragen opsporen met behulp van VAKIR en de onderwerpenlijst. Alhoewel op elke vraag een antwoord aanwezig was, waren de proefpersonen daarvan niet op de hoogte.

Tabel 2 geeft de resultaten van de proef:

	percentage correct beantwoorde vragen	zoektijd (s)
VAKIR	83	127
onderwerpenlijst	67	155

Tabel 2 : Resultaten per zoekmethode

Geconcludeerd kan worden dat zoeken met VAKIR sneller verloopt en vaker tot het juiste antwoord leidt dan zoeken met de onderwerpenlijst.

Er is onderzocht waarom in sommige gevallen de gewenste informatie niet wordt gevonden. Tabel 3 geeft de resultaten van dit onderzoek:

	keuze trefwoord	keuze informatieleverancier	keuze in informatie informatie-leverancier	andere
VAKIR	73	22	5	0
onderwerpenlijst	70	21	6	3

Tabel 3 : Oorsprong fouten in procenten

Het blijkt dat het vinden van het juiste trefwoord voor beide zoekmethoden het grootste probleem vormt. Een trefwoord wordt als onjuist beschouwd als het irrelevant is voor de te beantwoorden vraag, of niet is gerelateerd aan de pagina's met de gewenste informatie. Bij VAKIR is vooral het eerste trefwoord van belang. Vanwege het feit dat

de geselecteerde pagina's betrekking hebben op alle ingevoerde trefwoorden wordt de gewenste informatie direct onbereikbaar bij een onjuist eerste trefwoord. In 59 procent van de gevallen waarin een vraag onjuist werd beantwoord bleek dit het gevolg van een onjuist eerste trefwoord.

Keuze van de verkeerde informatieleverancier is vaak te wijten aan een verkeerde omschrijving van de inhoud van de pagina door de descriptor.

De derde oorzaak (keuze in informatie informatieleverancier) is het gevolg van slecht gestructureerde informatie van die informatieleverancier.

Als proefpersonen halverwege een correcte zoekweg stopten dan is dit gerubriceerd onder de rubriek 'andere'.

Tenslotte is gebleken dat de expliciete semantische hulp geen nuttig onderdeel vormt van de hulpfaciliteiten. De reden hiervoor is dat de aangeboden trefwoorden de gebruiker vaak misleiden.

3. EXPERIMENTEREN MET EEN INDEXEERPROGRAMMA

3.1 De indexeerproblematiek

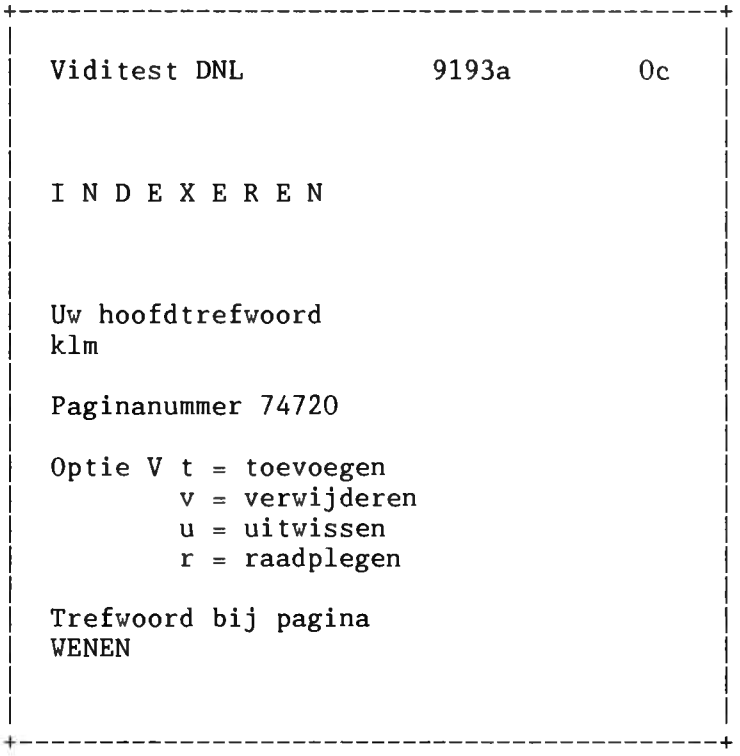
Voor het functioneren van VAKIR is het noodzakelijk dat de informatieleverancier hoofd- en/of subtrefwoorden aan zijn pagina's toekent. Het toekennen van deze trefwoorden aan een pagina wordt indexeren genoemd. Door een groep van twintig tijdelijke werknemers zijn ongeveer 40.000 Viditestpagina's geïndexeerd.

De functie van de aan een pagina toegekende trefwoorden is tweeledig:

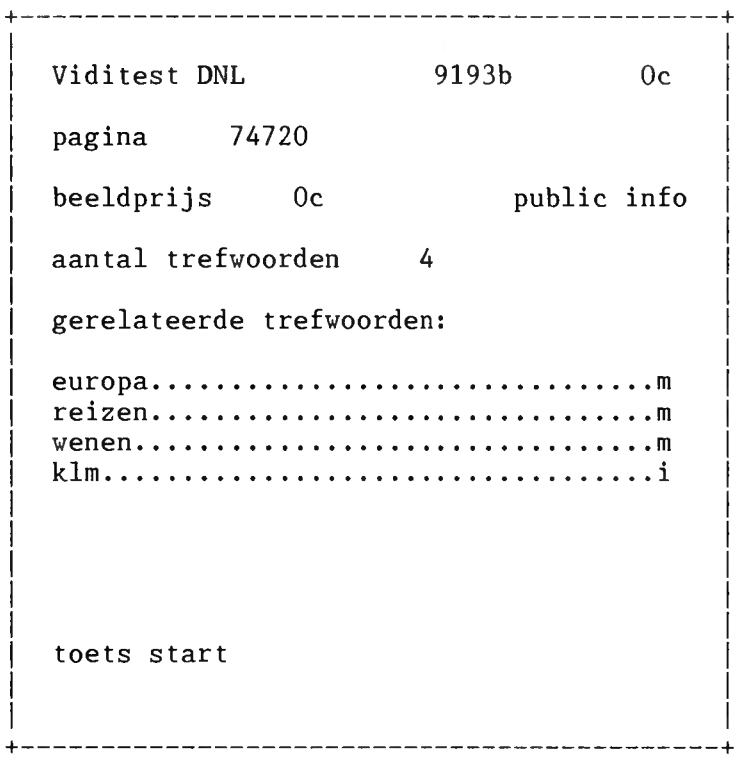
Ten eerste worden de toegekende trefwoorden gebruikt als zoekargument in fase één van de dialoog. Dit betekent dat op basis van de ingetoetste trefwoorden pagina's worden geselecteerd.

Ten tweede worden de toegekende trefwoorden gebruikt als descriptor in fase twee van de dialoog. Het trefwoord beschrijft de inhoud van een pagina in de lijst van geselecteerde pagina's.

Via een zogenaamde 'indexeerdialoog' kan de informatieleverancier ieder willekeurig hoofdtrefwoord of één van zijn subtrefwoorden aan een pagina toekennen, danwel van de pagina verwijderen. Daarvoor zijn de opties 't' en 'v' beschikbaar. Zie figuur 8 . Optie 'r' geeft een overzicht van alle aan de pagina toegekende trefwoorden, samen met verdere bijzonderheden over de pagina. Van elk trefwoord is het type gegeven: hoofdtrefwoord (m), informatieleveranciersnaam (i) of subtrefwoord (s). Zie figuur 9 . Optie 'u', voor het verwijderen van alle aan de pagina toegekende trefwoorden, is nog niet geïmplementeerd.



Figuur 8 : De bestaande indexeerdialoog



Figuur 9 : 'Raadplegen'

Bij het indexeren moet de informatieleverancier trachten die hoofd- en subtrefwoorden aan een pagina toe te kennen die door een willekeurige gebruiker als zoekargument voor de informatie op die pagina kunnen worden gebruikt. Het valt op dat de indexeerdialoog de informatieleverancier daar nauwelijks ondersteuning bij biedt.

Wellicht is het mogelijk om uit de tekst van de pagina zelf 'zinvolle' trefwoorden af te leiden. Om de mogelijkheden daartoe te onderzoeken is een indexeerprogramma geschreven dat op automatische wijze trefwoorden uit een pagina afleidt. Het ontwerp van dit programma wordt in de volgende paragraaf gegeven. Met behulp van dit programma is een experiment uitgevoerd waarin voor een aantal pagina's op automatische wijze trefwoorden zijn gegenereerd. De resultaten van dit experiment en de conclusies daaruit worden in paragraaf 3.3 gegeven.

3.2 Het ontwerp van een experimenteel indexeerprogramma

De informatieleverancier staat voor de taak om trefwoorden aan zijn pagina's toe te kennen. Om hem hierbij ondersteuning te kunnen bieden is onderzoek verricht naar de mogelijkheden om uit de tekst van een pagina trefwoorden af te leiden. Er zijn twee strategieën volgens welke trefwoorden uit een pagina kunnen worden afgeleid:

1. semantische strategie: op basis van een betekenisanalyse van de op een pagina voorkomende tekst worden geschikte trefwoorden bepaald.

2. lexicale strategie: spoor de in de pagina voorkomende hoofdtrefwoorden op.

De semantische strategie vereist de ontwikkeling van een ontleedprogramma waarmee de tekst op een pagina door een computer kan worden 'begrepen'. Het ontwikkelen van een dergelijk programma kost vele manjaren. Dit is de reden waarom uitwerking van de semantische strategie binnen dit afstudeerproject niet mogelijk is.

De lexicale strategie is relatief eenvoudig uit te werken. Hierbij dient van elk in de pagina gevonden woord te worden bepaald of dit een hoofdtrefwoord is of niet. Dit kan eenvoudig door te onderzoeken of het woord in de STL voorkomt.

Om de bruikbaarheid van de lexicale strategie te kunnen beoordelen is een experimenteel indexeerprogramma ontwikkeld. Dit programma genereert voor iedere Viditestpagina twee verzamelingen trefwoorden. De ene verzameling bestaat uit op basis van de lexicale strategie gevonden hoofdtrefwoorden, de andere bestaat uit de handmatig aan de pagina toegekende hoofd- en subtrefwoorden. Door het vergelijken van beide verzamelingen kan een indruk worden verkregen over de bruikbaarheid van de via de lexicale strategie gevonden trefwoorden.

Elke pagina bestaat uit één of meer beelden. Een beeld bevat een vaste hoeveelheid informatie, die precies op een beeldscherm kan worden getoond. Het programma analyseert het zogenaamde "a-beeld" van de pagina. Het a-beeld is het eerste beeld van de pagina. Alle in het a-beeld van de pagina voorkomende woorden worden opgespoord. Van ieder gevonden woord wordt bepaald of het een hoofdtrefwoord is of niet. Dit gebeurt door te onderzoeken of het in de STL voorkomt. Bij het zoeken in de STL wordt gebruik gemaakt van de impliciete morfologische hulpfaciliteiten van VAKIR. Als het woord in de STL blijkt voor te komen en nog niet eerder in de pagina is gevonden, dan is een nieuw trefwoord voor deze pagina bekend.

3.3 Resultaten uit een experiment met het indexeerprogramma

Met behulp van het indexeerprogramma is een 63-tal pagina's geanalyseerd. Deze 63 pagina's zijn ten behoeve van het experiment uit paragraaf 2.5 door een aantal experts geïndexeerd. Per pagina zijn twee verzamelingen trefwoorden gegenereerd. De ene verzameling bestaat uit de door de experts aan de pagina toegekende hoofd- en subtrefwoorden. Deze verzameling noemen we 'M'. De andere verzameling bevat de hoofdtrefwoorden die met behulp van de lexicale strategie zijn gevonden. Deze verzameling noemen we 'N'.

Van elke verzameling is het gemiddelde aantal trefwoorden bepaald. Het gemiddeld aantal trefwoorden in verzameling M noemen we 'm' en het gemiddelde aantal trefwoorden in N noemen we 'n'. Verder is het gemiddeld aantal trefwoorden bepaald dat M en N gemeenschappelijk hebben. Dit aantal noemen we 'p'. Tabel 4 geeft de gevonden waarden van m,n en p.

m:	6.2
n:	10.3
p:	2.0

Tabel 4

Bij het beoordelen van deze resultaten is aangenomen dat de trefwoorden uit verzameling M de inhoud van de pagina volledig en juist omschrijven. Alhoewel dit niet het geval hoeft te zijn, zijn de trefwoorden uit verzameling M het enige vergelijkingmateriaal dat voor handen is. Met deze aanname is het mogelijk de bruikbaarheid van de lexicale strategie te beoordelen. De trefwoorden uit verzameling M worden de voor deze pagina relevante trefwoorden genoemd. Ieder ander trefwoord is per definitie een voor deze pagina irrelevant trefwoord. Er is een tweetal kwaliteitscriteria voor de trefwoorden uit verzameling N gedefinieerd. Deze zijn:

1. de 'volledigheid' v .
   Deze is gedefinieerd als:
   $v = p/m$, met $0 \leq v \leq 1$ en $m \neq 0$

   De volledigheid geeft de fractie van het aantal toegekende trefwoorden dat met behulp van de lexicale strategie is gevonden. Als v=1 dan zijn alle toegekende trefwoorden met de lexicale strategie gevonden.

2. de 'juistheid' j .
   Deze is gedefinieerd als:
   $j = p/n$, met $0 \leq j \leq 1$ en $n \neq 0$

De juistheid geeft de fractie van het aantal met de lexicale strategie gevonden trefwoorden dat aan de pagina is toegekend. Als j=1 dan is elk met behulp van de lexicale strategie gevonden trefwoord aan de pagina toegekend.

In het ideale geval is verzameling N gelijk aan M. In dat geval geldt: v=1 en j=1.

Uit de gegevens van tabel 4 kunnen v en j worden bepaald. Deze blijken te zijn:

v:	.32
j:	.19

Tabel 5

Het blijkt dat zowel v als j aanzienlijk kleiner zijn dan 1. Wat is daarvan de betekenis voor het indexeren?

Een lage waarde van v betekent dat slechts een gering deel van de relevante trefwoorden met behulp van de lexicale strategie is gevonden. Indien in de verzameling toegekende trefwoorden relevante trefwoorden ontbreken dan betekent dit een verslechterde bereikbaarheid van de informatie op die pagina. Als namelijk een gebruiker de informatie op die pagina tracht te selecteren via een ontbrekend relevant trefwoord dan heeft dit tot gevolg dat de betreffende pagina wordt 'weggeselecteerd'. Voor de gebruiker lijkt het of de gewenste informatie niet aanwezig is.

Een lage waarde van j betekent dat van alle gevonden trefwoorden een groot deel irrelevant is. Het toekennen van irrelevante trefwoorden houdt eveneens een verslechterde toegankelijkheid van de informatie in. In dit geval kunnen bij het opvragen van informatie een aantal pagina's ten onrechte worden geselecteerd. Dit maakt het voor de gebruiker lastiger om de juiste pagina's te vinden.

Bovenstaande resultaten zijn besproken met een ergonoom, die betrokken is geweest bij de experimenten uit paragraaf 2.5 . Er is een analyse gemaakt van de inhoud van alle 63 pagina's. Dit leidde tot het bijstellen van de verzameling toegekende trefwoorden voor in totaal 28 pagina's. In de meeste gevallen betrof dit het toekennen van één of meer trefwoorden, die met behulp van de lexicale strategie voor de betreffende pagina waren gevonden. Verder zijn de woorden uit verzameling N die een synoniem bleken te zijn van een trefwoord uit verzameling M, alsnog tot de overeenkomende woorden gerekend.

Na het aanbrengen van deze wijzigingen zijn m, n en p opnieuw bepaald. Tabel 6 geeft hiervan de resultaten.

m:	7.7
n:	10.3
p:	3.3

Tabel 6

Uit deze gegevens kunnen v en j worden bepaald. Deze blijken te zijn:

v:	0.43
j:	0.32

Tabel 7

Het blijkt dat door de aangebrachte veranderingen zowel v als j is verbeterd.

Er is onderzocht door welke oorzaken niet alle aan de pagina toegekende trefwoorden met behulp van de lexicale strategie worden gevonden. Tabel 8 geeft de resultaten van dit onderzoek.

oorzaak	percentage gevallen
1) toegekend trefwoord komt niet in de pagina voor	42
2) toegekend trefwoord is een samengesteld trefwoord	25
3) toegekend trefwoord komt in de pagina voor als onderdeel van een samengesteld zelfstandig naamwoord	14
4) toegekend trefwoord bevindt zich op vervolgbeeld	8
5) falende enkelvoud-meervoud substitutie	6
6) overige	5

Tabel 8 : Foutoorzaken in procenten

Toelichting bij tabel 8 :

ad 2) Een samengesteld trefwoord is een trefwoord dat uit twee of meer afzonderlijke woorden bestaat. Bijvoorbeeld: abstracte kunst, middelbare technische scholen. Het indexeerprogramma herkent alleen enkelvoudige woorden.

ad 3) Een samengesteld zelfstandig naamwoord is een woord dat bestaat uit twee of meer woorden die elk ook afzonderlijk kunnen voorkomen. Bijvoorbeeld: postmuseum, autotest. Het indexeerprogramma herkent alleen het volledige woord en niet de afzonderlijke delen ervan.

ad 4) Door het programma wordt alleen het a-beeld geanalyseerd, zodat trefwoorden die op een vervolgbeeld voorkomen niet worden gevonden.

ad 5) VAKIR tracht een woord dat in de enkelvoudsvorm is gevonden naar de meervoudsvorm te transformeren. De regels die daarvoor worden toegepast leiden echter niet in alle gevallen tot de gewenste meervoudsvorm. Is dit laatste het geval dan zal de enkelvoudsvorm van een in de pagina gevonden trefwoord niet worden herkend.

Conclusie:
Via de lexicale strategie kan gemiddeld 43 procent van de voor een pagina relevante trefwoorden worden opgespoord. Van alle gevonden

trefwoorden blijkt gemiddeld 32 procent relevant te zijn. De gevonden trefwoorden kunnen aan de informatieleverancier worden aangeboden als suggestie voor aan zijn pagina toe te kennen trefwoorden. Het niet vinden van relevante trefwoorden is voornamelijk het gevolg van het ontbreken van deze trefwoorden in de tekst van de pagina.

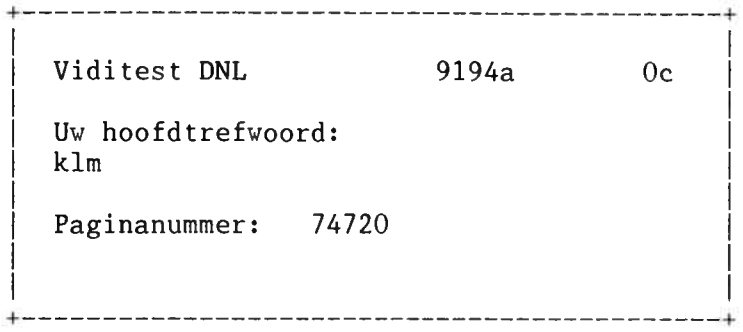
4. EEN NIEUWE INDEXEERDIALOOG

4.1 Het ontwerp van de nieuwe indexeerdialoog

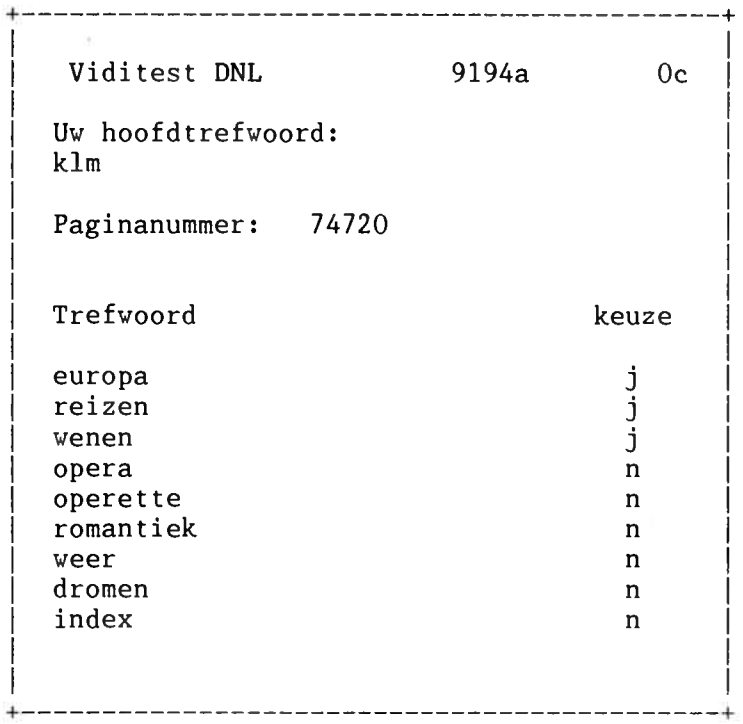
Het is mogelijk gebleken om uit de tekst van een pagina suggesties voor aan de pagina toe te kennen trefwoorden af te leiden. Omdat de huidige indexeerdialoog nauwelijks ondersteuning biedt bij het indexeren en zich slecht leent voor het aanbieden van suggesties is besloten een nieuwe indexeerdialoog te ontwikkelen.

De nieuwe dialoog vindt plaats op een zogenaamd dialoogbeeld. Bovenaan dit beeld verschijnt de naam van de informatieleverancier. Vervolgens kan de informatieleverancier het paginanummer van één van zijn pagina's intoetsen. Zie figuur 10a .

Na het intoetsen van het paginanummer verschijnt een lijstje met trefwoorden voor de gekozen pagina. Zie figuur 10b . Als dit lijstje niet op één dialoogbeeld past, dan kan de rest ervan op een vervolgbeeld worden bekeken.



Figuur 10a



Figuur 10b

Figuur 10a,b : de nieuwe indexeerdialoog

Het lijstje van trefwoorden bestaat uit twee delen:

Het eerste deel bestaat uit de hoofd- en subtrefwoorden die aan de pagina zijn toegekend. Indien er nog geen trefwoorden aan de pagina zijn toegekend, dan is dit deel van het lijstje leeg.

Het tweede deel van het lijstje bestaat uit alle in het a-beeld van de pagina gevonden hoofdtrefwoorden die niet aan de pagina zijn toegekend. Deze trefwoorden worden 'suggesties' genoemd.

Achter elk trefwoord is een keuzeveld aanwezig. Daarin kan worden aangegeven of het trefwoord aan de pagina moeten worden toegekend of niet. Invullen van een 'j' (ja) betekent dat het trefwoord moet worden toegekend en invullen van een 'n' (nee) betekent dat het trefwoord niet moet worden toegekend. De keuzevelden zijn door het systeem van een 'default'-keuze voorzien. Deze is 'j' voor de aan de pagina toegekende trefwoorden en 'n' voor de suggesties. De informatieleverancier kan zijn keuze maken uit de aangeboden trefwoorden door de keuzevelden van de gewenste keuze te voorzien.

Na het invullen van het laatste keuzeveld wordt om bevestiging van de aangebrachte veranderingen gevraagd, of kan, indien van toepassing, de de rest van het lijstje worden bekeken. Als bevestiging van de veranderingen plaats vindt dan worden deze uitgevoerd.

Op ieder moment kan het indexeren worden afgebroken door het toetsen van de <START>-toets. In dat geval worden er geen wijzigingen uitgevoerd.

Een speciale gebruiker van de dialoog is de beheerder van het trefwoordensysteem. Hij wordt in staat gesteld om bovenaan het dialoogbeeld de naam van een informatieleverancier in te voeren. Daarmee beschikt hij over de mogelijkheid om elke willekeurige pagina te indexeren.

4.2 Een mini-enquete over de nieuwe indexeerdialoog

Een drietal personen is geinterviewd over de nieuwe indexeerdialoog. Dit waren een informatieleverancier, een ergonoom en een technicus. De ergonoom en de technicus zijn beiden betrokken geweest bij de ontwikkeling van VAKIR. Aan alle drie is een demonstratie van de dialoog gegeven en zijn de door hen gemaakte opmerkingen genoteerd. De belangrijkste daarvan worden hieronder weergegeven.

- De informatieleverancier

  De informatieleverancier toont zich enthousiast over de dialoog en ziet deze als een welkome aanvulling. Door het aanbieden van suggesties bestaat volgens hem het gevaar dat informatieleveranciers deze klakkeloos overnemen. Dit kan leiden tot het toekennen van relatief veel irrelevante trefwoorden en daarmee de toegankelijkheid van de informatie verslechteren. De informatieleverancier pleit daarom voor het instellen van een 'indexeercommissie' die steekproefsgewijs en met een hoge pakkans controleert hoe de pagina's zijn geïndexeerd en zonodig correctief optreedt.

  Met de dialoog is het alleen mogelijk om trefwoorden uit het aangeboden lijstje toe te kennen. Het moet echter mogelijk

zijn om elk willekeurig hoofd- of subtrefwoord toe te kennen. Daarom moet de dialoog worden uitgebreid met de mogelijkheid om zogenaamde 'vrije keuzen' in te voeren.

- De ergonoom

  De ergonoom vindt het opsporen van suggesties uit de tekst van de pagina een interresante en zinvolle nieuwe mogelijkheid. De dialoog behoeft aanpassing. Dit betreft een duidelijkere scheiding tussen de aan de pagina toegekende trefwoorden en de suggesties. Verder moet op elk gewenst moment de mogelijkheid bestaan om vrije keuzen in te voeren. Tenslotte moet, indien de informatieleverancier een suggestie accepteert of een toegekend trefwoord wil verwijderen, dit trefwoord direct worden toegekend danwel worden verwijderd. De wijziging wordt zichtbaar gemaakt door het verplaatsen van het trefwoord van het lijstje met suggesties naar het lijstje met toegekende trefwoorden of andersom.

- De technicus

  Ook de technicus ziet de nieuwe dialoog als een welkome aanvulling. Evenals de anderen merkt hij op dat de mogelijkheid tot vrije keuzen moet worden geboden.

5. AANBEVELINGEN

5.1 Toevoegen van de mogelijkheid tot vrije keuzen

Met de nieuwe indexeerdialoog is het alleen mogelijk om trefwoorden uit het aangeboden lijstje aan de pagina toe te kennen. Het moet echter mogelijk zijn om elk willekeurig hoofd- of subtrefwoord toe te kennen. Daarom dient de dialoog te worden uitgebreid met de mogelijkheid om zogenaamde 'vrije keuzen' in te voeren.

5.2 Duidelijkere melding over aanwezigheid vervolgbeeld

Het lijstje van trefwoorden kan de lengte van één beeld overschrijden. De informatieleverancier wordt daar pas op attent gemaakt na het invullen van het laatste keuzeveld in het eerste deel van het lijstje. Beter is om daarvan direct melding te maken, bijvoorbeeld door het geven van de boodschap 'vervolgbeeld aanwezig'. De informatieleverancier kan zo eerst kennis nemen van alle trefwoorden en vervolgens daaruit zijn keuze maken.

6. VERSLAG VAN DE BEOORDELINGSZITTING

Dit hoofdstuk bevat een verslag van de beoordelingszitting zoals deze plaats vond op 25 augustus 1986. De examencommissie bestond uit de volgende 4 leden:

- Professor ir. J.L. de Kroes (TH-Delft).
- ir. D. Sparreboom (TH-Delft).
- ing. N.C.J.M. de Beer (DNL).
- dr. ir. L.P.A.S. van Noorden (Afdeling Sociaal Wetenschappelijk onderzoek PTT).

van Noorden: In paragraaf 3.3 wordt gesproken over een 63-tal pagina's dat is geanalyseerd. Totaal zijn er 40.000 pagina's geindexeerd. Waarom zijn daarvan slechts 63 geanalyseerd en waarom juist deze?

van der Lippe: Van het totaal van 40.000 pagina's zijn die 63 pagina's door experts geindexeerd en de overige door een aantal vakantiewerkers. Omdat niet kan worden ingestaan voor de kwaliteit van door vakantiewerkers toegekende trefwoorden vormen deze geen bruikbaar vergelijkingsmateriaal bij het beoordelen van de resultaten van de lexicale strategie.

van Noorden: Je bent vooral geinteresseerd in de mening van de gewone gebruiker en daar zijn de door vakantiewerkers geindexeerde pagina's wellicht wel bruikbaar voor.

van Noorden: Een van de oorzaken van het niet vinden van trefwoorden blijken de zogenaamde samengestelde zelfstandige naamwoorden te zijn. Is het zinvol ook deze in de STL op te nemen?

van der Lippe: Er is voor gekozen samengestelde zelfstandige naamwoorden niet in de STL op te nemen om zodoende het aantal trefwoorden in de STL te beperken. Samengestelde zelfstandige naamwoorden worden dan omschreven via het intoetsen van hun afzonderlijke delen.

van Noorden: Wie waren de experts uit paragraaf 2.5?

van der Lippe: Mij is alleen bekend dat één van hen een ergonoom was.

de Beer: Tabel 8 vermeldt een aantal foutoorzaken. Een daarvan is het niet analyseren van eventuele vervolgbeelden. Is het niet zinvol ook

deze te analyseren?

van der Lippe: Het analyseren van vervolgbeelden zal leiden tot het vinden van meer relevante trefwoorden maar betekent tevens een forse stijging van de verwerkingstijden bij het indexeren. De reden daarvoor is dat ook alle woorden uit de vervolgbeelden in de STL moeten worden opgezocht.

Sparreboom: Indexeren vindt minder frequent plaats dan het zoeken zodat de stijging van de verwerkingstijden van minder belang is.

de Beer: Informatie in Viditel is geordend in een bepaalde hiërarchische structuur. Is het geen zinvolle aanbeveling om ook uit de menupagina's van de bestaande zoekmethoden trefwoorden af te leiden voor de onderliggende pagina's?

van der Lippe: Dit kan inderdaad een zinvolle aanbeveling zijn. Echter wanneer VAKIR in Viditel wordt geïntroduceerd en succesvol blijkt te zijn, bestaat de mogelijkheid dat VAKIR de bestaande zoekmethoden verdringt. Bovenstaande aanbeveling is dan minder bruikbaar.

de Beer: Zijn de bestaande zoekmethoden en VAKIR volledig substitueerbaar?

van der Lippe: Alhoewel beiden niet geheel substitueerbaar zijn, bestaat het vermoeden dat VAKIR de bestaande zoekmethoden op den duur gaat verdringen. Dit omdat is gebleken dat zoeken met de bestaande zoekmethoden moeizaam verloopt en VAKIR de informatie aanzienlijk beter toegankelijk maakt.

de Beer: De heer de Beer poneert een stelling die van der Lippe naar keuze kan verdedigen of aanvallen. De stelling luidt:

"Het is aan te bevelen een pagina altijd automatisch, dat wil zeggen zonder enige interventie van de informatieleverancier te indexeren met suggesties zoals die door het indexeerprogramma worden aangeboden".

Van der Lippe besluit de stelling aan te vallen.

van der Lippe: Het blijkt dat via de lexicale strategie gemiddeld 43 procent van de voor een pagina relevante trefwoorden wordt gevonden. Als de pagina's alleen automatisch worden geïndexeerd dan zal een groot deel van relevante trefwoorden niet worden toegekend. Dit leidt tot een verslechterde toegankelijkheid van de informatie op die pagina.

de Beer: De stelling sluit niet uit dat gemiste relevante trefwoorden niet op een andere manier alsnog worden toegekend.

van der Lippe: In dat geval worden er aan de pagina relatief veel irrelevante trefwoorden gekoppeld.

de Kroes: Wellicht is van de via de lexicale strategie gevonden

irrelevante trefwoorden een gedeelte niet echt irrelevant en zou daarom ook aan de pagina kunnen worden gekoppeld. Dit zou een argument zijn voor de stelling. De irrelevante trefwoorden kunnen wellicht worden opgedeeld in twee groepen: zinvolle en niet-zinvolle trefwoorden.

de Beer: Is het toekennen van een enkel irrelevant trefwoord een nadeel?

van der Lippe: Nee, het toekennen van een enkel irrelevant trefwoord is geen nadeel. Een gebruiker kan een pagina ten onrechte selecteren via een irrelevant trefwoord. Als echter de gebruiker besluit om nog een extra trefwoord in te toetsen dan is de kans groot dat dit niet aan die pagina is toegekend. De ten onrechte geselecteerde pagina wordt dan weggeselecteerd.

de Beer: Van de gevonden niet-zinvolle trefwoorden is wellicht een gedeelte eenvoudig uit te sluiten.

van der Lippe: Als dit mogelijk is is de stelling verdedigbaar, anders vooralsnog niet.

Sparreboom: Een andere definitie van de volledigheid en juistheid is mogelijk door deze voor iedere pagina apart te bepalen en dan te middelen over het aantal pagina's. Over deze mogelijkheid vind ik niets in het verslag. Is deze ook overwogen?

van der Lippe: In een gesprek met de Beer is deze mogelijkheid inderdaad bestudeerd maar daarvan heb ik ten onrechte niets in het verslag vermeld.

Sparreboom: Volgens mij klopt de berekening van de volledigheid uit paragraaf 3.3 niet. Deze moet 0.43 zijn.

van der Lippe: Dit zal ik na rekenen en zonodig veranderen.

Sparreboom: De resultaten van de onderwerpenlijst uit tabel 2 zijn aanzienlijk beter dan die uit tabel 1. Hoe is dit te verklaren?

van der Lippe: Aan de onderwerpenlijst zijn tussentijds verbeteringen aangebracht. Daarnaast betreft het steekproeven die wellicht voor enige spreiding in de uitkomsten zorgen.

van Noorden: Uit paragraaf 4.1 blijkt dat er ook subtrefwoorden aan een pagina toegekend kunnen zijn. Subtrefwoorden kunnen toch niet met de nieuwe dialoog worden gekoppeld?

van der Lippe: Deze zijn met behulp van de oude dialoog toegekend.

de Kroes: Er mist een lijst van gebruikte afkortingen.

van der Lippe: Deze zal ik toevoegen.

de Kroes: De literatuurlijst is wel erg kort. Is er niet meer literatuur bestudeerd?

van der Lippe: Ik heb wel meer literatuur gelezen maar die is niet daadwerkelijk voor het verslag gebruikt.

de Beer: Voor zover mij bekend is dit de enige literatuur die over dit onderwerp bestaat.

van Noorden: In de mini-enquete ontbreekt het oordeel van een professionele indexeerder.

van der Lippe: Deze heb ik inderdaad niet geinterviewd.

de Kroes: Bestaat er geen mechanisme om weinig gebruikte trefwoorden uit het bestand te verwijderen?

van der Lippe: Dit is momenteel niet geautomatiseerd. De systeembeheerder is daarvoor verantwoordelijk.

BIJLAGE A

LITERATUURLIJST

1. Velthoven, R.H. van
Zoekgedrag bij Viditel,
een laboratorium-onderzoek.
SWO-verslag 616/12
DNL-verslag 482 ST
Dr. Neher Laboratorium Leidschendam, juli 1981

Een laboratorium-evaluatie van de gebruikersvriendelijkheid van bestaande zoekmethoden in Viditel.

2. Zuidam, H.
A keyword search method for interactive videotex
DNL-REPORT 519 ST/84
Dr. Neher Laboratorium Leidschendam, juli 1984

Resultaten van een onderzoek naar een gebruikersvriendelijke zoekmethode voor interactieve videotex.

3. Weerdmeester, B.A.
User experiment with a keyword search method for Viditel
SWO-REPORT 616/20
DNL-REPORT 513 ST/84
Dr. Neher Laboratorium Leidschendam, augustus 1984

Een laboratoriumexperiment waarin een nieuw ontwikkelde trefwoordenzoekmethode (VAKIR) wordt vergeleken met één van de bestaande zoekmethoden.