

***DIE INVLOED VAN MENSLIKE INLIGTINGS-
VERWERKING OP DIE KWALITEIT VAN
INDEKSERING***

Christine Nel, B. Bibl.

Skripsie voorgelê vir gedeelte nakoming van die vereistes vir die graad Magister Bibliothecologiae in die Departement Inligtingstudies aan die Potchefstroomse Universiteit vir Christelike Hoër Onderwys.

Leier: Mev. M A Buys
Potchefstroom
1995

Abstract

In order to retrieve information successfully, the quality of indexing and indexing terms must be good. Two methods of indexing are discussed, i.e. assigned and derivative indexing. In assigned indexing the subject content of the documents is analysed by the indexer. After concepts have been identified, the best ones are selected. Consequently a number of mental processes (i.e. human information processing) are present. Indexing terms are abstracted from the title or text by means of a computer program. The process of subject analysis in assigned indexing is also in accordance with the user enquiring process.

To ascertain the quality of indexing and indexing terms, a process of evaluation is necessary. Different parameters can be used in evaluation. The measuring units giving the best indication of the quality of indexing are retrieval and precision. Recall determines the number of relevant documents being retrieved, while precision shows the non-relevant documents being held back. In an *ex post facto* experiment conducted it has been found that assigned indexing terms give the best results regarding retrieval and precision.

The conclusion can be made that assigned indexing with its element of human information processing yields indexing terms of a better quality than derivative indexing.

INHOUDSOPGAWE

BLADSY

HOOFSTUK 1	1
INLEIDING	1
1.1 Agtergrond	1
1.2 Probleemstelling	2
1.3. Doelstellings	4
1.4. Metode van ondersoek	4
1.5. Verdere hoofstukindeling	5
HOOFSTUK 2	6
TOEGEKENDE EN AFGELEIDE INDEKSERING	
2.1 Inleiding	6
2.2 Die proses van indeksering	7
2.3 Die proses by toegekende indeksering	11
2.4 Indekseertaal	13
2.4.1 Definiëring van 'n indekseertaal	13
2.4.2 Soorte indekseertale	14
2.4.2.1 Beheerde indekseertaal (gekontroleerde indekseertaal)	14
2.4.2.2 Natuurlike indekseertaal	15
2.4.2.3 Vrye indekseertaal	15
2.4.3 Toegekende terme as beheerde indekseertaal	16
2.5 Afgeleide indeksering	17
2.6 Verskille tussen afgeleide en toegekende indeksering	19
2.7 Voorbeelde van afgeleide indeksering	19
2.7.1 Titelaafgeleide indeksering	20
2.7.2 Koördinasie van woorde in die titel	20
2.7.3 Woordfrekwensie	21
2.7.4 Linguistiese analise van taal deur die rekenaar	22

HOOFSTUK 3

DIE KOGNITIEWE PROSES TYDENS MENSLIKE INLIGTINGSVERWERKING EN DIE ROL DAARVAN BY INDEKSERING

3.1 Inleiding	24
3.2 Die kognitiewe proses	24
3.2.1 Fisiologiese samestelling van die menslike brein	24
3.2.2 Natuurlike funksionering van die menslike geheue	25
3.3 Menslike inligtingsverwerking	26
3.4 Die Van Dijk-Kintsch teorie oor die interpretasie van taal	29
3.5 Stappe in die indekseerproses	34
3.5.1 Analisering van die dokument en die vastelling van die onderwerpsinhoud	35
3.5.2 Begrip van die dokumentinhoud	36
3.5.3 Identifisering van konsepte	37
3.5.4 Die keuse van konsepte	37
3.5.5 Oordrag van konsepte na die indekseertaal	38
3.6. Die sockproses	40
3.6.1 Stappe in die sockproses	42
3.6.2 Ooreenkoms tussen die indekseerproses en die sockproses	44

HOOFSTUK 4

DIE KWALITEIT VAN INDEKSERING EN INDEKSEERTERME

4.1. Inleiding	47
4.2 Faktore wat die prestasie van die stelsel beïnvloed	48
4.2.1 Akkuraatheid van indeksering	48
4.2.2 Uitputtendheid van indeksering	49
4.2.3 Eenvormigheid by die keuse van terme	50
4.3 Evaluering van die prestasievermoë van 'n IOHS	52
4.4 Herwinning en presiesheid as meeteenhede	54
4.5 Die invloed van die indekseerterme op herwinning en presiesheid	58
4.5.1 Die invloed van spesifiekheid van terme	58
4.5.2 Die invloed van 'n beheerde woordeskat	60

HOOFSTUK 5	62
'N EVALUERING VAN TOEGEKENDE INDEKSERING TEENoor AFGELEIDE INDEKSERING	63
5.1 Inleiding en probleemstelling	63
5.2 Navorsingsontwerp	65
5.3 Seleksie van databasisse	65
5.3.1. Ferdinand Postma Biblioteek se intydse katalogus	66
5.3.2 In U Lig-databasis	66
5.3.3 Repertorium van Suid-Afrikaanse tydskrifartikels	66
5.3.4 ERIC	67
5.4 Metode van ondersoek	67
5.5 Sockresultate	69
5.5.1 Resultate wat die herwiningsprestasie van titelafgeleide en toegekende indeksering aandui	69
5.5.2 Resultate wat die presiesheid van herwinde dokumente aandui	71
5.6 Analise van die resultate	79
5.6.1 Herwinning	79
5.6.2 Relevansie (presiesheid)	79
5.6.2.1 Intydse katalogus	80
5.6.2.2 ERIC	80
5.6.2.3. Repertorium	80
5.6.2.4. In U Lig	80
HOOFSTUK 6	82
AFLEIDINGS EN GEVOLGTREKKING	82
6.1 Inleiding	82
6.2 Afleidings wat gemaak kan word	82
6.3. Die rol van 'n beheerde woordeskat	84
6.4. Gevolttrekking	85
6.5 Slotopmerking	86
BIBLIOGRAFIE	87

LYS VAN FIGURE

Figuur 1 Fases van 'n inligtingverwerkingsmodel	27
Figuur 2 Kommunikasie tussen die gebruiker en die bibliotekaris (volgens Shannon & Weaver)	42
Figuur 3 Ooreenkoms tussen die indekseerproses en die soekproses	44

DANKBETUIGINGS

- My studieleier, Mev. M.A. Buys, vir uitstaande leiding en die deeglikheid waarmee sy my in hierdie studie gelei het.
- My werkgewer, Prof. C.J.H. Lessing, Hoof van die Departement Inligtingstudies, wat hierdie studie moontlik gemaak het deur die nodige fasiliteite en leiding te verskaf.
- Dr. C.S. Reinecke, vir die uitstekende taalversorging en al die moeite wat sy gedoen het.
- Ernst Drewes, wat my baie bygestaan en ondersteun het tydens dié studie asook die bydrae wat hy gelwer het tot die redegeryng van die teks.
- My ouers, Pieter en Albertha Nel, sonder wie se volgehoue aanmoeding hierdie studie nie moontlik sou wees nie en die finansiële bystand wat hulle verleen het tot die voltooiing van my studie.

*“For the Lord is great,
and greatly to be praised;” (Ps. 96:4).*

HOOFSTUK 1

INLEIDING

1.1 Agtergrond

Inligtingsontsluiting is 'n belangrike voorvereiste vir die latere effektiewe herwinning van inligting. Gedokumenteerde inligting, dit wil sê inligting wat in een of ander fisiese medium opgeteken is, kan net vir gebruikers van nut wees as hulle van die bestaan van sulke inligting weet en waar dit gevind kan word. Sodanige kontrole oor gepubliseerde inligting word deur middel van bibliografiese beheer uitgeoefen. Laasgenoemde hou onder andere in dat die inhoud van dokumente effektief geïndekseer moet word, dit wil sê, volgens onderwerpsinhoud ontsluit moet word. Met die oog op inligtingsherwinning moet die indekseerterme wat by ontsluiting gebruik word, aanvaarbaar en bruikbaar wees, veral vanuit 'n gebruikersoogpunt. Al die indekseerterme vorm saam die indekseertaal. Boon (1979:6) omskryf 'n indekseertaal as 'n kunsmatige tussentaal wat daarop ingestel is om kommunikasie tussen die gebruikers van inligting en die inhoud van inligtingsbronne te bewerkstellig. Die wyse waarvolgens die meeste biblioteekgebruikers na inligting soek, is volgens die onderwerpsbenadering; daarom is die indekseertaal 'n belangrike komponent van die inligtingsontsluiting- en -herwinningstelsel. Effektiewe indeksering behels nie net besluitneming oor die inhoud van 'n dokument en hoe om die inhoud deur middel van indekseerterme uit te druk nie, maar dit gaan ook oor die belang wat gebruikers by 'n spesifieke dokument kan hê, naamlik om te verseker dat hulle by dié dokument sal uitkom wat hulle navraag reg sal beantwoord.

Indekseerterme kan volgens twee metodes vasgestel word. Volgens die eerste metode analiseer die indekseerder die inhoud en identifiseer hy/sy 'n aantal konsepte wat die onderwerpsinhoud van die dokument verteenwoordig. Uit hierdie konsepte word 'n paar gekies wat die onderwerpsinhoud die beste uitdruk. Met behulp van 'n gestandaardiseerde lys terme word die konsepte wat gekies is, na indekseerterme omgeskakel. Dié metode staan bekend as toegekende indeksering. Menslike inligtingsverwerking ("human information processing") is dus hierby betrokke.

Volgens die tweede metode wat deur Foskett (1986:7) verduidelik word, word woorde wat in die titel of teks van die dokument voorkom, netso as indekseerterme gebruik. Geskikte woorde om as indekseerterme te gebruik, word byvoorbeeld gekies op grond van die frekwensie waarvolgens hulle in die teks voorkom, sintaktiese patrone in 'n sin, ensovoorts. Hierdie metode is bekend as afgeleide indeksering en menslike logika speel geen rol hierin nie, aangesien hierdie metode uitsluitlik deur 'n rekenaar uitgevoer word.

Verder sê Lancaster (1986:1) dat die proses wat tydens inligtingsherwinning plaasvind, netso kompleks is as dié wat tydens onderwerpsanalise realiseer en dat dit ook met menslike logika ooreenstem. Tydens herwinning word die navraag op dieselfde wyse geanaliseer as wat die geval met onderwerpsanalise tydens indeksering is. Om die verlangde inligting op te spoor, word konsepte na soekterme wat met die indekseerterme moet ooreenstem, omgeskakel. Hierdie proses stem grootliks ooreen met die menslike verstandspeser wat tydens toegekende indeksering plaasvind.

1.2 Probleemstelling

Hutchins (1978:172-173) wys daarop dat min navorsing nog oor die eerste metode van indeksering gedoen is:

"We find a great deal about the construction of index languages and classification systems, about the principles of classification, about the correct formulation of index entries (e.g. the uses of standard citation orders and of chain indexing) and about the evaluation of indexes and information systems. But we find little about how indexers and classifiers decide what the subject of a document is, how they decide what it is 'about'."

Oor die kognitiewe proses wat tydens indeksering plaasvind, is min navorsing gedoen, asook oor die invloed wat dit op die kwaliteit van indeksering het. Todd (1992:101) sluit by Hutchins se standpunt aan as hy sê:

"... published accounts of indexing seem to ignore or skate over the mental processes that take place during subject analysis. There is little about how indexers decide what the subject of a document is, how they determine its aboutness. More interest is currently being shown in automatic indexing than

in improving human indexing. This lack of focus on the cognitive processes of indexing - in discussion and in research - has contributed to a number of indexing problems that continue to hound the profession."

Volgens Wellisch (1989:9) is daar duidelik twee gedagterigtings oor hierdie indekseermetodes in die literatuur op te merk. Die voorstaanders van afgeleide indeksering is van mening dat alle gebruikersnavrae doeltreffend beantwoord kan word sonder enige inagneming van die rol wat die verstandspeser van die mens tydens indeksering speel. Volgens Wellisch (1989:9) sien hulle dit as "a brave new world in which all or most questions will be dealt with by machines, which will also execute the subsequent search and retrieval of information unfailingly and to the full satisfaction of the inquirer"

Die ander gedagterigting beklemtoon dat daar nie sonder meer aangeneem kan word dat bogenoemde rekenaarmatige indekseermetodes goed genoeg sal presteer om gebruikersnavrae bevredigend te beantwoord nie. Volgens Dyson (soos aangehaal deur Wellisch, 1988:10) word inligting deur die mens vir die mens geskep en is dit steeds die menslike element wat aan die einde van die hele inligtingsoordragproses tevrede gestel moet word. Soos reeds gesê, word die beste konsepte tydens onderwerpsanalise deur 'n indekseerder gekies om die onderwerp uit te druk. Hierdie proses van menslike inligtingsverwerking kan nie deur 'n rekenaar nagedoen word nie. Gevolglik word die sin of betekenis van woorde nie by afgeleide indeksering in berekening gebring nie. Dit is verder opmerklik dat groot indekseringsdienste soos ERIC, Medline en vele ander, steeds van toegekende indeksering in hulle databasisse gebruik maak.

In die lig van bogenoemde kan die volgende vrae gestel word:

- Wat behels die proses van toegekende indeksering en hoe verskil dit van afgeleide indeksering?
- Watter bydrae lewer menslike inligtingsverwerking tot die kwaliteit van indekseerterme?
- Hoe presteer toegekende indeksering in vergelyking met afgeleide indeksering tydens inligtingsherwinning?

1.3. Doelstellings

Daar word met hierdie studie beoog om:

- die prosesse by toegekende en afgeleide indeksering te ontleed en die verskille aan te dui;
- vas te stel watter rol menslike inligtingsverwerking in die toekenning van indekseerterme speel en hoe dit die kwaliteit van terme beïnvloed;
- die prestasie van toegekende en afgeleide indeksering met betrekking tot herwinning en presiesheid te vergelyk.

1.4. Metode van ondersoek

1.4.1 Literatuurstudie

'n Literatuurstudie is gedoen om die prosesse van toegekende en afgeleide indeksering te ontleed en om die kognitiewe proses wat tydens die lees, verstaan en begryp in die menslike brein plaasvind, te beskryf. Verder is die literatuur nagegaan om vas te stel wat die eienskappe van indekseertaal is en om 'n aantal meeteenhede vir evaluering te ondersoek. Nog doelstellings was om die invloed van indekseerterme op herwinning en presiesheid te bestudeer, om die stappe in die soekproses te beskryf en ten slotte om ooreenkomste en verskille tussen indeksering en die soekproses aan te dui.

1.4.2 Eksperimentele ondersoek

'n Ex post facto-eksperiment is uitgevoer om die prestasievermoë van toegekende en afgeleide indeksering aan die hand van voorbeelde met mekaar te vergelyk. Dit is gedoen deur 'n aantal tydskrifartikels wat oor sekere vakterreine handel en wat deur albei indekseermetodes ontsluit word, in soektogte te gebruik om vas te stel wat die herwinning en presiesheid van elke metode is. Soekterme wat gebruik is, kan óf as toegekende óf as afgeleide terme funksioneer. Soekresultate is beoordeel met die oog op herwinningsprestasie en presiesheid (relevansie van herwinde dokumente).

1.5, Verdere hoofstukindeling

In Hoofstuk 2 sal die proses van indeksering bespreek word met die doel om die verskille tussen die twee indekseermetodes aan te dui. Indekseertaal en die eienskappe daarvan word aan die hand van die drie soorte indekseertale verduidelik, naamlik 'n beheerde woordeskat, natuurlike indekseertaal en vrye indekseertaal. Afgeleide en toegekende indeksering word aan die hand van voorbeelde bespreek.

In Hoofstuk 3 sal die kognitiewe proses wat tydens menslike inligtingsverwerking plaasvind en die rol daarvan by indeksering bespreek word. Verder sal aangetoon word hoe die Van Dijk-Kintsch teorie oor die interpretasie van taal op die stappe in die indekseerproses toegepas kan word, terwyl die soekproses en die ooreenkoms en verskil tussen die indekseer- en die soekproses ook ondersoek sal word.

In Hoofstuk 4 sal die kwaliteit van indeksering en indekseerterme aan die hand van faktore wat die prestasie van die stelsel kan beïnvloed, bespreek word. Herwinning en presiesheid sal as meeteenhede by evaluering ondersoek word, terwyl die belangrikheid van relevansie en pertinensie vir gebruikersverwagtinge ook aangetoon sal word.

In Hoofstuk 5 sal die prestasie van toegekende en afgeleide indeksering aan die hand van voorbeelde met mekaar vergelyk word. Die resultate wat verkry word, sal geanaliseer word om te bepaal wat die herwinning en presiesheid van elke metode is.

In Hoofstuk 6 word afleidings oor die prestasievermoë van toegekende en afgeleide indeksering gemaak, voordat tot 'n gevolgtrekking oor die rol wat menslike inligtingsverwerking by die kwaliteit van indeksering speel, gekom word.

HOOFSTUK 2

TOEGEKENDE EN AFGELEIDE INDEKSERING

2.1 Inleiding

In hierdie hoofstuk word twee metodes van indeksering wat radikaal van mekaar verskil, bestudeer. In die eerste metode, naamlik toegekende indeksering, speel menslike inligtingsverwerking 'n deurslaggewende rol, terwyl dit in die ander metode, naamlik afgeleide indeksering, glad nie ter sprake is nie. Voordat daar egter na hierdie twee indekseermetodes gekyk word, is dit eers nodig om aandag te gee aan indeksering in die algemeen en hoe dit plaasvind.

Indeksering word soos volg in die BDI (1990:142) omskryf:

- "1. Het analyseren van de inhoud van een publikatie op de meest karakteristieke elementen en het vertalen van die elementen in termen en structuur van een indexsysteem;
- "2. het maken van een register."

UNESCO (1975:2) omskryf indeksering as "the fact of describing and identifying a document in terms of its subject content". Rowley (1988:48) beskou dit as "an attempt to present a summary of document content " en gaan voort, "because index terms must be used as access points, the summarization of document content achieved in indexing documents must be more tightly structured." Feinberg (1973:1) se omskrywing lui: "Indexing consists of indicating the subject content of an item of information by assigning one or more terms to the document so as to categorise it."

Uit bogenoemde omskrywings is daar gemeenskaplike begrippe wat as die basiese kenmerke van indeksering geïdentifiseer kan word:

- dit is die proses waarvolgens die onderwerpskenmerke van 'n dokument ontleed en geïdentifiseer word;

- hierdie onderwerpskenmerke word gekondenseer ("summary") en in indekseerterme uitgedruk;
- die indekseerterme vorm soekpunte in katalogusse, indekse of ander bibliografiese hulpmiddele.

'n Aantal doelwitte van indeksering kan ook uit bogenoemde punte afgelei word. As eerste doelwit kan gestel word dat dokumente volgens hul onderwerpsinhoud ontsluit moet word met die oog op die daarstel van bibliografiese hulpmiddele (soos katalogusse en indekse) wat later gebruik kan word om die inligting te herwin. Wanneer bogenoemde hulpmiddele by herwinning gebruik word, byvoorbeeld om 'n aantal verwante dokumente te identifiseer, of slegs 'n gedeelte van 'n dokument wat met 'n navraag verband hou, word die navraag op dieselfde wyse as die vasstelling van indekseerterme gehanteer. Die navraag word in individuele konsepte ontleed en na die terme wat in die indekseertaal gebruik word, omgeskakel (UNESCO, 1975:2).

'n Verdere doelwit is dat die indekseerterme soekpunte in die indeks vorm en daardeur die gebruiker na die inligting moet lei. Volgens Feinberg (1973:1) is 'n indeks "a bridge between the contents of the literature and the user." Dit moet dus as brug tussen die gebruiker en die dokument dien en in hierdie proses speel die indekseerterme (indekseertaal) 'n belangrike rol, soos later in die studie sal blyk. Die indekseerterme is 'n handige en kort verwysingsstelsel om dokumentinhoud in die ontsluitingstelsel te hanteer. Dit bepaal of die vloeï van inligting glad en sonder probleme kan verloop.

Om toegekende indeksering as metode van onderwerpsontsluiting te kan bespreek, is dit nodig om vooraf kortliks na die stappe in die indekseerproses te kyk. Dit word in 3.6 vollediger bespreek om aan te toon hoe menslike inligtingsverwerking plaasvind.

2.2 Die prosesse van indeksering

In teenstelling met katalogisering wat volgens gestandaardiseerde reëls gedoen word, is daar geen vaste reëls wat by indeksering gebruik kan word

nie. Verskillende instansies het wel riglyne opgestel waarvolgens die proses kan plaasvind. So 'n riglyn is dié van UNESCO, bekend as die *UNISIST indexing principles*, en sal hier as voorbeeld gebruik word om die proses van indeksering te verduidelik. Hiervolgens bestaan die indekseerproses uit die volgende twee basiese stappe (UNESCO, 1975:3):

Eerste stap: Vasstelling van die onderwerpsinhoud

In die vasstelling van die onderwerpsinhoud word die volgende substadia onderskei:

- 'n begrip van die onderwerpsinhoud van 'n dokument en die doel van die outeur daarmee;
- die identifisering van konsepte wat die onderwerpsinhoud uitdruk;
- die keuse van konsepte wat die onderwerpsinhoud die beste uitdruk.

Tydens die indekseerproses neig hierdie drie substadia om te oorvleuel.

Tweede stap: Omskakeling van konsepte na indekseerterme

In die omskakeling van konsepte na indekseerterme kom die indekseertaal ter sprake. In 2.4 sal aangetoon word hoe indekseerterme vasgestel word. Daar sal ook later aangedui word watter invloed die gebruik van gestandaardiseerde terme op herwinning en presiesheid het.

Ander outeurs se uiteensetting van die indekseerproses stem in 'n groot mate ooreen met die proses soos deur UNESCO uiteengesit. Volgens Vickery (1968:355) is dit die proses waarin inligting in so 'n mate ontleed word dat 'n aantal kernwoorde in die dokument geïdentifiseer kan word. Op hulle beurt druk hulle die onderwerp kernagtig uit. Die konsepte wat gekies word, moet gebruik word om relevante dokumente te identifiseer. Verder moet hulle soekpunte verskaf wat in 'n literatuursoektog gebruik kan word om die aard

en inhoud van 'n dokument weer te gee of om as 'n substituuat vir 'n dokument te dien.

Lancaster (1991:8) beskryf die verskillende stappe in die indekseerproses soos volg:

- konseptuele ontleding en
- omskakeling van konsepte ooreenkomstig die indekseertaal.

Alhoewel Lancaster die twee stappe van mekaar onderskei, kan hulle nie altyd presies afgebaken word nie en kán hulle ook gelyktydig plaasvind. Konseptuele ontleding behels dat die indekseerder moet besluit waarom 'n dokument handel. Omdat indeksering daarop gemik is om aan die behoeftes van gebruikers te voldoen, is doeltreffende indeksering nie net 'n besluitnemingsproses om te bepaal waarom 'n dokument handel nie, maar die behoeftes van gebruikers moet ook in ag geneem word, sowel as die redes waarom die dokument vir 'n sekere groep gebruikers van belang is. Volgens Lancaster (1991:8) kan daar nie sommer net aangeneem word dat daar slegs een korrekte stel indekseerterme is nie. 'n Dokument kan verskillend deur verskillende instansies geïndekseer word en behóórt ook verskillend geïndekseer te word wanneer meer as een groep gebruikers daarin geïnteresseerd sou wees.

'n Begrip wat in hierdie verband ter sake is, is die sogenaamde "aboutness" of *waaroor* van 'n dokument. Lancaster (1991:10) bevraagteken die noodsaaklikheid om die "aboutness" van 'n dokument te verstaan voordat effektief geïndekseer kan word. Hy is van mening dat dit genoeg is om te besef dat 'n dokument vir 'n sekere groep gebruikers van belang is. Volgens hom behoort hierdie besef aan die indekseerder die nodige insig te verskaf om te bepaal watter konsepte vir indeksering van belang is. Hierdie siening behoort die proses van konseptuele ontleding te vereenvoudig, terwyl die oordrag van hierdie konsepte na die indekseertaal wat gebruik word, 'n besluit sal wees wat net daarop toegespits is om te bepaal watter woorde (indekseerterme) die beste daarby pas. Tinker (1966:97) se standpunt is egter dat 'n indekseerder by die toekenning van 'n indekseerterm seker moet maak dat die term relevant vir die dokument is: ". . . that is, he asserts that the

meaning of the descriptor is strongly associated with a concept embodied in the document, and that it is appropriate for the subject area of the document."

Wat die stappe in die indekseeerproses betref, verdeel Cleveland & Cleveland (1990:104-105) hulle soos volg:

- onderwerpsanalise en
- onderwerpsbepaling.

Onderwerpsanalise word volgens hulle deur verskillende faktore beïnvloed. Ten eerste word die keuse van konsepte deur die indekseerbeleid wat gevolg word, bepaal. Laasgenoemde kan byvoorbeeld voorskrifte bevat ten opsigte van die hoeveelheid tyd wat afgestaan word om 'n dokument te bestudeer - dokumente kan oorsigtelik of intensief ondersoek word. Ook die diepte van indeksering en die keuse van konsepte sal 'n rol speel.

Ten tweede is hulle van mening dat die keuse van konsepte reeds tydens onderwerpsanalise plaasvind. Die indekseerder besluit d n reeds watter aspekte van die onderwerp beklemtoon moet word en watter nie. Hierdie konsepte word uit die volgende dele van 'n dokument gekies: die titel, die ekserp, die teks self en die bibliografie.

Rowley (1987:174-175) se uiteensetting van die stappe wat by indeksering gevolg word, stem grootliks ooreen met di  wat deur Cleveland & Cleveland voorgestel word. Hulle behels die volgende:

- Die indekseerder moet met die onderwerpsinhoud van die dokument vertrou raak. Sy noem dit "familiarisation". Konsepte word gelyktydig uit spesifieke gedeeltes van die teks, naamlik die titel, inhoudsopgawe, hoofstukindelings, ekserpte en inleiding verkry.
- In die tweede stap onderskryf Rowley die standpunt van Cleveland & Cleveland dat net konsepte wat die moeite werd is, geidentifiseer behoort te word. Dit word teen die agtergrond van die spesifieke indekseeromgewing waarin gewerk word, gedoen.
- Die laaste stap is wanneer besluit word watter aspekte van 'n onderwerp na die terme van die indekseertaal omgeskakel moet word.

Chu & O'Brien (1993:439) beklemtoon net soos Cleveland & Cleveland en Lancaster dat onderwerpsanalise die belangrikste stap in indeksering is. Onderwerpsontleding in 'n dokument moet eers plaasvind alvorens die terme wat gekies is, na die van die indekseerstelsel wat gebruik word, omgeskakel kan word. Die indekseerproses bestaan volgens Chu & O'Brien uit dieselfde stappe as die wat deur UNESCO beskryf word.

Uit al die voorafgaande uiteensettings van die indekseerproses, kan twee gemeenskaplike fasette geïdentifiseer word, naamlik:

- onderwerpsanalise vind plaas;
- die geskikste indekseerterme word deur die indekseerder gekies.

Hierdie twee aspekte oefen 'n belangrike invloed op die kwaliteit van indeksering uit, soos later in 5.6.2 aangetoon sal word.

Volgens Lancaster (1991:13) behels die oordrag van konsepte na indekseerterme "the conversion of the conceptual analysis of the document into a particular set of index terms." Hy tref 'n onderskeid tussen "indexing by extraction" (afgeleide indeksering) en "indexing by assignment" (indeksering deur toekenning). In afgeleide indeksering word woorde of frases wat in die dokument voorkom, netso gebruik om die onderwerpsinhoud te verteenwoordig. By toegekende indeksering word terme toegeken wat uit 'n ander bron as die dokument self afkomstig is, naamlik 'n aparte lys terme wat gestandaardiseer is. Dit staan bekend as die indekseertaal en sal in 2.4 breedvoeriger bespreek word.

Vervolgens word die proses van toegekende indeksering ondersoek. Afgeleide indeksering kom in 2.5 aan die beurt.

2.3 Die proses by toegekende indeksering

Die proses wat tydens toegekende indeksering plaasvind, bestaan uit al die stappe wat in 2.2 volgens die UNISIST-beginsels uiteengesit is.

Tydens die proses van toegekende indeksering word die inhoud van 'n dokument deur die indekseerder ontleed en 'n besluit word geneem oor watter indekseerterme geskik is om gebruik te word. Hier speel menslike redenasie ("human information processing") 'n deurslaggewende rol. Volgens Foskett (1986:68) staan toegekende indeksering ook as konsepindeksering bekend. 'n Konsep is volgens hom 'n "idea of a class of objects; general notion" - met ander woorde, dit is die begrip of idee wat 'n mens van iets het. 'n Konsep is nie noodwendig net een woord nie. Daar is enkelvoudige konsepte, byvoorbeeld *planeete*, en ook saamgestelde konsepte, byvoorbeeld *lewe op planeete*. Met die toekenning van indekseerterme behoort die konsep uitgedruk word, of dit nou enkelvoudig of saamgesteld is. Foskett (1986:68-70) wys verder daarop dat die geskikste konsepte, net soos by die UNISIST-beginsels, aan die hand van 'n skema of hiërargie gekies word.

In toegekende indeksering kan die onderwerp van 'n dokument volledig ontsluit word, wat beteken dat daar óf vir elke faset of subfaset van die onderwerp 'n indekseerterm toegeken word, óf terme kan ook selektief gekies word, wat impliseer dat slegs 'n beperkte aantal terme gekies word om die sentrale onderwerpsterrein van 'n dokument te ontsluit. Hoe meer terme toegeken word, hoe toegankliker is die dokument en hoe beter sal dit herwin kan word. Verder behoort terme ook so spesifiek moontlik gekies te word. Lancaster (1991:26) beskou die spesifiekheid van indekseerterme as die belangrikste beginsel van onderwerpsontsluiting. Die aspek sal vollediger in 2.4 behandel word.

Lancaster (1991:221), asook Cleveland & Cleveland (1990:7), is van mening dat alle menslike indeksering toegekende indeksering is. Die indekseerder maak staat op sy eie agtergrondskennis van 'n onderwerp en die begrip wat hy van die betekenis van 'n woord het, sal sy keuse van terme beïnvloed. Verder sal 'n indekseerder beïnvloed word deur die aantal kere wat die terme in die dokument voorkom en ook deur die plek waar hulle voorkom, soos in die titel, opsomming, onderskrifte of by illustrasies. In toegekende indeksering moet 'n indekseerder eers 'n groot aantal dokumente indekseer voordat hy/sy oor 'n grondige kennis van 'n onderwerp sal beskik. Konsepte wat te spesifiek gekies word, het soms tot gevolg dat 'n konsep verskillende betekenis in verskillende dissiplines het. Die konsep *basis* sal byvoorbeeld

heeltmal 'n ander betekenis vir 'n generaal in die weermag as vir 'n basketbalspeler hê.

Vervolgens word eers na die indekseertaal gekyk, omdat dit die hulpmiddel/apparaat is wat in toegekende indeksering gebruik word.

2.4 Indekseertaal

2.4.1 Definiëring van 'n indekseertaal

Volgens Boon (1979:6) is 'n indekseertaal 'n kunsmatige tussentaal wat daarop ingestel is om kommunikasie tussen die gebruikers van inligting en die inhoud van inligtingsbronne te bewerkstellig.

Pao (1989:102) definieer indekseertaal as die totale versameling indekseerterme wat in 'n stelsel vir indeksering gebruik word. Daarteenoor sê Cleveland & Cleveland (1990:78) dat 'n indekseertaal 'n vasgestelde versameling van natuurlike woorde is wat toelaatbaar is om 'n spesifieke dokument te beskryf. Hierdie indekseerterme word nie net gebruik om 'n dokument te beskryf nie, maar hulle lei die gebruiker ook na relevante dokumente deurdat hulle as soekpunte in bibliografiese rekords in 'n katalogus of indeks gebruik word. Verder wys Cleveland & Cleveland daarop dat indekseertaal ook gebruik word om 'n navraag te formuleer en nie net om verwysings na relevante dokumente te verskaf nie. Ander outeurs soos Aluri *et al.* (1991:28) en Harter (1986:22) se omskrywings van 'n indekseertaal stem grotendeels hiermee ooreen.

Uit al die bogenoemde omskrywings kan die volgende aspekte as belangrike eienskappe van 'n indekseertaal geïdentifiseer word:

- die indekseertaal moet as tussentaal kommunikasie tussen die dokument en die gebruiker bevorder;
- indekseerterme moet die onderwerpsinhoud van dokumente uitdruk;
- indekseerterme vorm soekpunte in 'n indeks;

- 'n gebruiker se navraag word ooreenkomstig die indekseertaal geformuleer.

Daar sal in 2.4.2 en 3.3 aangetoon word wat die verband van elk van bogenoemde punte met toegekende en afgeleide indeksering is, asook hoe menslike inligtingsverwerking 'n invloed daarop uitoefen.

Inligtingsherwinning (net soos inligtingsontsluiting) is altyd 'n kommunikatiewe proses (Harter, 1986:22) en in hierdie proses speel die indekseertaal 'n primêre rol. Daarom is dit nodig om in meer besonderhede na sekere aspekte van indekseertaal te kyk.

2.4.2 Soorte indekseertale

Daar is verskillende soorte indekseertale. Rowley (1987:169-170) onderskei drie basiese tipes, naamlik beheerde, natuurlike en vrye indekseertaal.

2.4.2.1 Beheerde indekseertaal (gekontroleerde indekseertaal)

Volgens Rowley (1987:168-169) word terme in beheerde indekseertaal uit 'n gestandaardiseerde lys gekies. Dit beteken dat nie enige woord as 'n indekseerterm gebruik kan word nie, maar dat daar op 'n voorkeurterm besluit word wat dan konsekwent gebruik word. Uit die terme *huwelik*, *bruilof* en *troue* kan *huwelik* byvoorbeeld as voorkeurterm gekies word. Die gebruik van een voorkeurterm hou in dat kruisverwysings vanaf die ander terme gemaak moet word. Twee tipes gekontroleerde indekseertale kom voor, naamlik alfabetiese indekseertaal en klassifikasieskemas. Alfabetiese indekseertaal word in ontsluitingshulpmiddele soos tesourusse en onderwerpshoofde (saakhoofde) gevind. Kontrole word deurgaans oor hierdie terme uitgeoefen deur kruisverwysings vanaf variante terme te maak wat nie as deskriptore of saakhoofde gebruik word nie. Daarteenoor word in klassifikasieskemas aan elke onderwerp 'n notasie toegeken wat die onderwerp in konteks plaas (gewoonlik op 'n hiërgiesse wyse). Vir die doel

van hierdie bespreking sal die klem hoofsaaklik op verbale indekseertaal val, naamlik deskriptore en saakhoofde.

2.4.2.2 Natuurlike indekseertaal

Natuurlike indekseertaal is nie 'n selfstandige taal wat apart soos 'n beheerde indekseertaal funksioneer nie. Enige woord wat in die teks voorkom, kan as 'n moontlike indekseerterm gebruik word. In die praktyk hou dit in dat die indekseerder van terme gebruik maak wat in die teks of titel van 'n dokument voorkom. Hy hoef dus nie terme uit 'n gestandaardiseerde lys te kies of hulle daarin te kontroleer nie.

Hierdie tipe indeksering word vandag meestal gebruik waar indeksering deur die rekenaar gedoen word en in al hierdie metodes word geen menslike oordeel of logika gebruik wat die keuse van terme betref nie. Anders as by beheerde indekseertaal, kan natuurlike indekseertaal deur die mens sowel as 'n rekenaar uitgevoer word. Indien daar van 'n rekenaar gebruik gemaak word, word terme op grond van 'n spesifieke program deur die rekenaar "gekies".

2.4.2.3 Vrye indekseertaal

Volgens Rowley (1987:169-170) is vrye indekseertaal nie 'n afsonderlike indekseertaal wat uit spesifieke terme bestaan nie. Die terme verskil ook nie van die terme wat gebruik word om konsepte in die literatuur uit te druk nie. Indeksering is vry in die sin dat daar geen beperking geplaas word op die terme wat gebruik mag word nie. Dit beteken dat woorde uit die teks of enige ander bron (byvoorbeeld toegekende terme soos saakhoofde) gebruik kan word. Vrye indekseertaal verskil van natuurlike indekseertaal in die sin dat natuurlike indekseertaal beperk word tot die woordeskat van die dokument wat geïndekseer word. Vrye indekseertaal word geensins beperk nie en kan enige term insluit. Wanneer daar van 'n menslike inset gebruik gemaak word, is vrye indekseertaal afhanklik van die kennis en vaardigheid van die

indekseerder. Vryetaal-indeksering met behulp van 'n rekenaar is vir alle praktiese doeleindes dieselfde as natuurliketaal-indeksering.

2.4.3 Toegekende terme as beheerde indekseertaal

Soos reeds genoem, is saakhoofde en deskriptore voorbeelde van beheerde indekseertaal. Beheerde indekseertaal is die resultaat of produk van toegekende indeksering. By die toeken van saakhoofde of deskriptore is die hele proses van menslike inligtingsverwerking soos wat in Hoofstuk 3 bespreek sal word, teenwoordig. Nadat onderwerpsanalise plaasgevind het, word die geskikste konsepte uitgekies en omgeskakel na of vertaal in indekseerterme. Hierdie omskakeling vind plaas aan die hand van 'n gestandaardiseerde lys terme. Dit verseker dat dieselfde term konsekwent vir 'n bepaalde konsep gebruik word.

Saakhoofde kan as ware beheerde of gekontroleerde woordeskat gesien word, want dit bevat volgens Pao (1989:119) "predetermined authorized terms with elaborate syntactic rules for application". Met 'n saakhoof word probeer om die hele onderwerpsinhoud van 'n dokument uit te druk. Dit kan wissel van enkelvoudige tot saamgestelde terme waar die "syntactic rules" 'n belangrike rol speel. By deskriptore is die terme dikwels enkelvoudiger as wat die geval by saakhoofde is. Die rede hiervoor is dat die dokumentinhoud tydens onderwerpsanalise in enkelvoudige konsepte ingedeel en deur enkelvoudige terme uitgedruk word. In die soekfase word terme gekoördineer om 'n saamgestelde onderwerp te soek (met ander woorde postgekoördineerde terme) (Pao,1989:118):

"The searcher analyzes the sought topic into its component concepts, identifies their corresponding index terms, and then synthesizes them by coordinating the terms into a search statement representing the desired topic."

Deskriptore is egter ook terme wat uit 'n gestandaardiseerde lys gekies is, dit wil sê hulle word gekies nadat onderwerpsanalise plaasgevind het en nie uit die titel of teks afgelei nie.

In 'n beheerde woordeskat word ook verwante terme aangedui, byvoorbeeld breër of enger terme. Hierdeur word addisionele terme gesuggereer wat dalk vir 'n bepaalde konsep toepasliker kan wees.

Woordeskatbeheer

Die waarde van woordeskatbeheer is dat dit die beste of aanvaarbaarste terme vir die indekseerder voorstel. 'n Beheerde woordeskat stel die indekseerder in staat om tussen terme wat dieselfde idee verteenwoordig, te onderskei. Verder dra 'n beheerde woordeskat ook daartoe by dat die terme wat die konsep die beste beskryf, gekies kan word. 'n Goed gekontroleerde woordeskat help die indekseerder om tot volle begrip van 'n dokument se onderwerp te kom deur terme voor te stel wat die indekseerder lei tot die mees spesifieke, akkuraatste en toepaslikste term om die konsep mee uit te druk. 'n Goed beheerde woordeskat dra verder daartoe by dat die gebruiker ook tydens die soekproses by die regte term uitkom om hom na die dokument te lei. Vir die gebruiker verskaf woordeskatbeheer waardevolle leidrade vir die formulering van 'n soekstelling en die herwinning van die maksimum relevante dokumente.

Samevatting

In bogaande bespreking is aangetoon dat die proses van toegekende indeksering menslike redenasie en oordeel insluit. Die analise van dokumentinhoud, die identifisering van konsepte, die keuse van die geskikste konsepte en die uitdruk daarvan in indekseertaal, behels 'n groot hoeveelheid verstandspesesse.

2.5 Afgeleide indeksering

Vervolgens sal gekyk word hoe afgeleide indeksering plaasvind en hoe dit van toegekende indeksering verskil. Volgens Bourne (1963:10) is die

eenvoudigste metode van indeksering om woorde uit die dokument self as indekseerterme te gebruik. Hierdie woorde kan uit die titel of teks van die dokument of selfs uit 'n ekserp onttrek word. Dié metode van indeksering word meestal deur 'n rekenaarprogram uitgevoer. Enige betekenisvolle woord in die dokument kan deur die rekenaar as indekseerterm onttrek word. Niebetekenisvolle woorde (byvoorbeeld voorsetsels, lidwoorde en voornaamwoorde) word as moontlike indekseerterme uitgeskakel deur van 'n stopwoordlys gebruik te maak.

Volgens Cleveland & Cleveland (1990:79) word afgeleide indeksering ook natuurlike of vrye indekseertaal genoem (soos reeds verduidelik in 2.4.2.2 en 2.4.2.3). Afgeleide indeksering is gegrond op die feit dat dieselfde woord wat deur verskillende outeurs gebruik word om 'n onderwerp te beskryf, as indekseerterm gebruik kan word. Afgeleide indeksering streef dus daarna om direk met die leser op 'n eenvoudige, logiese wyse te kommunikeer deur terme te gebruik wat in die natuurlike taal van die dokument voorkom. Cleveland & Cleveland merk op dat toegekende indeksering in vergelyking met afgeleide indeksering 'n geforseerde, kunsmatige taal is wat tot gevolg het dat daar onnodige kommunikasieversperrings tussen 'n outeur en leser ontstaan.

Aluri *et al.* (1991:98) meld verder dat afgeleide indeksering ook vanuit enige onderwerpsdraende veld in 'n bibliografiese rekord gedoen kan word, byvoorbeeld uit:

- die titel, reeksname, konferensies en vergaderings;
- inhoudsaantekeninge.

Afgesien van bogenoemde kan ander velde ook bruikbare indekseerterme verskaf om onderwerpstoegang te bewerkstellig, naamlik:

- aanhalings;
- ekserpte;
- ongekontroleerde woordeskatte wat uit die indekse agter in boeke afgelei word;

- indekseerterme wat deur 'n outeur voorgestel word.

Hoewel hierdie velde nie altyd volledig by die bibliografiese rekord ingesluit is nie, moet die bruikbaarheid daarvan vir inligtingsherwinning in gedagte gehou word.

2.6 Verskille tussen afgeleide en toegekende indeksering

Hoewel sekere kriteria op albei metodes van indeksering toegepas word, verskil afgeleide indeksering van toegekende indeksering in die opsig dat die inhoud van 'n dokument nie ontleed word nie. Daar vind dus nie onderwerpsanalise soos in 2.2 bespreek is, plaas nie en geen besluit word geneem oor watter indekseerterme geskik is om gebruik te word nie.

In afgeleide indeksering word die verwantskap tussen indekseerterme nie gegee nie, wat wel die geval by toegekende indeksering is. Daar is geen woordeskatbeheer nie. Die gevolg hiervan is dat dieselfde konsep deur verskillende terme uitgedruk kan word, byvoorbeeld *rough*, *uneven*, *coarse*, *crude* (Aluri *et al.*, 1991:36). Dit hou bepaalde gevolge vir herwinning in. As 'n gebruiker nie aan alle moontlike sinonieme dink waaronder hy kan soek nie, sal hy nie alle dokumente oor 'n onderwerp herwin nie.

Vervolgens word 'n paar voorbeelde van afgeleide indeksering kortliks ondersoek.

2.7 Voorbeelde van afgeleide indeksering

In die volgende paar voorbeelde van afgeleide indeksering kan gesien word dat dit deur 'n rekenaarprogram uitgevoer word en dat menslike inligtingsverwerking nie teenwoordig is nie.

2.7.1 Titelaafgeleide indeksering

As 'n voorbeeld van titelaafgeleide indeksering, word KWIC (*Key Word in Context*) kortliks bespreek. 'n KWIC-indeks bestaan uit 'n lys titels. Elke betekenisvolle woord wat in 'n titel voorkom, word as 'n toegangspunt (indekseerterm) gebruik. Die sleutelwoord word in die middel van 'n bladsy geplaas, terwyl die oorblywende woorde in die titel links en regs van die sleutelwoord gerangskik word. 'n KWIC-indeks is die eenvoudigste vorm van afgeleide indeksering. Die waarde van so 'n indeks is dat elke sleutelwoord in sy konteks bestudeer kan word, dit wil sê soos dit in die titel voorkom. KWIC-indekse dui normaalweg 'n dokumentnommer aan wat benodig word om die volledige bibliografiese rekord in 'n meesterlêer na te speur. Die rekenaarprogram wat die indeks saamstel, identifiseer sleutelwoorde volgens 'n "omgekeerde" prosedure. Hierdie program is so opgestel dat stopwoorde, dit wil sê die woorde wat nie betekenisvol is nie en dus nie as sleutelwoorde kan funksioneer nie, nie as toegangspunte gebruik word nie. Die woorde in die stopwoordlys het wel 'n sintaktiese funksie, byvoorbeeld lidwoorde, voorvoegsels, agtervoegsels en voegwoorde, maar hulle is nie ekspressief ten opsigte van die onderwerp nie (Lancaster, 1991:45-47).

Hier vind geen menslike inligtingsverwerking plaas nie en die keuse van indekseerterme is heeltemal afhanklik van woorde wat in die titel voorkom. Foskett (1986:38) sê dat die titel dié deel van 'n dokument is waar die outeur die onderwerp gewoonlik probeer definieer. In sommige gevalle is dit 'n duidelike beskrywing van die inhoud, maar die omgekeerde geld ook. Soms gebeur dit dat die titel glad nie beskrywend van die onderwerp is nie, of slegs een faset daarvan aandui. KWIC-indeksering kan ook uit ander dele van die teks afgelei word, byvoorbeeld uit ekserpte (Lancaster, 1991:45-47).

2.7.2 Koördinasie van woorde in die titel

As voorbeeld van die koördinasie van woorde in die titel word permuterm-indeksering (*Permuted Terms*) genoem. Dieselfde prosedure wat met titelaafgeleide indeksering gevolg word, geld hier, maar twee betekenisvolle woorde uit die titel word op 'n keer gebruik en nie die hele titel nie. Hierdie

twee woorde word as 'n hoof en subhoof in die indeks gedruk om 'n termpaar te vorm. Dis makliker leesbaar as KWIC, maar gee nie die hele konteks weer nie. Elke sleutelwoord in die titel word op sy beurt met elkeen van die ander sleutelwoorde van die titel in die indeks gekoördineer.

So word alle moontlike verwante assosiasies van terme wat met 'n sekere konsep verband hou, in pare vertoon. Elke sleutelwoord word dan as 'n inskrywingspunt in die indeks gebruik (Lancaster, 1991:48). Bourne (1963:18) en Foskett (1986:38) sê dat die bruikbaarheid van hierdie tipe indekse grootliks bepaal word deur hoe kernagtig 'n dokumenttitel die inhoud van 'n dokument beskryf. Titelverryking kan gebruik word om titels ekspressiewer te maak.

2.7.3 Woordfrekwensie

In indeksering wat op woordfrekwensie berus, word indekseerterme gekies op grond van die aantal kere wat dit in 'n teks voorkom. H.P Luhn het 'n statistiese metode ontwikkel waarvolgens die rekenaar woorde in 'n teks tel en dan die betekenisvolheid van woorde volgens hulle frekwensie en rangorde in die teks bepaal. Daar word ook van twee afsnyppunte gebruik gemaak. Woorde wat bokant die boonste afsnyppunt voorkom, is te algemeen en word uitgelaat. Woorde wat onder die onderste afsnyppunt voorkom, is te seldsaam en word ook uitgelaat. Die beste indekseerterme lê halfpad tussen die twee afsnyppunte (Cleveland & Cleveland, 1990:227-228).

Volgens Cleveland & Cleveland is Luhn se metode daarop gebaseer dat woorde wat die meeste in die teks herhaal word, die hoofidees van die teks reflekteer. In hierdie metode lei die rekenaar indekseerterme op grond van hul frekwensie uit die teks af. Geen menslike inligtingsverwerking vind hier plaas nie.

2.7.4 Linguistiese analise van taal deur die rekenaar

Met linguistiese analise van die taal deur die rekenaar word probeer om die die rekenaar die teks te laat "verstaan" sodat geskikte indekseerterme gekies kan word. 'n Voorbeeld hiervan is FASIT (*Fully Automatic Syntactically based Indexing of Text*).

In FASIT word alle woorde in die teks in sintaktiese kategorieë verdeel. Hierdie kategorieë is gebaseer op die rededele van die Engelse taal, byvoorbeeld selfstandige naamwoorde, byvoeglike naamwoorde, werkwoorde, voorsetsels. Verder word formules volgens Engelse taalreëls opgestel en volgens hierdie formules word frases of woorde deur die rekenaar gekies wat as moontlike indekseerterme gebruik kan word. Aan hierdie woorde of frases word gewigte toegeken wat op hulle frekwensie in die teks gebaseer is. Uiteindelik word slegs daardie frases/woorde wat die grootste gewig dra, deur die rekenaar as indekseerterme geselekteer (Dillon & McDonald, 1983:101-102).

Die rekenaar word dus gebruik om sinvolle terme in die teks te identifiseer deur die sintaktiese patroon van laasgenoemde vas te stel. Met hierdie indekseermetode word probeer om die teks met behulp van die rekenaar te "interpreteer" sodat geskikte indekseerterme gekies kan word (Dillon & Gray, 1983:99-101).

Slotopmerking

Milstead (1984:134-135) wys op die volgende aspekte by afgeleide indeksering:

- Die ontwerper van 'n afgeleide indekseerstelsel moet steeds besluit of die subjektiewe of objektiewe betekenis van woorde in die keuse van woorde gebruik moet word. Die tipe terme wat deur die rekenaar gekies word, moet vooraf deur 'n mens gespesifiseer word, byvoorbeeld of negatiewe woorde of positiewe woorde, of 'n kombinasie hiervan, gebruik gaan word.

- Subhoofde in 'n indeks behoort eerder volgens sintaktiese verwantskap tussen terme of blote permutasie vasgestel te word en nie op grond van toevallige woordnabyheid van die woorde in die sin of titel nie.
- In afgeleide indeksering word die gebruik van arbeidsintensiewe onderwerpsontleding van 'n dokument vermy, wat tot gevolg het dat indeksering baie vinniger en ekonomieser kan plaasvind. Anders as in die geval van toegekende indeksering waar woordbetekenis deur die verskaffing van breër en enger terme uitgebrei word, word woordbetekenis in afgeleide indeksering beperk. Om 'n hiërargie van verwante terme in afgeleide indeksering saam te stel, is ingewikkeld. Die rede is dat terme op verskillende maniere afgelei word en daar geen sprake van enige woordeskatbeheer is nie. Alle tipes afgeleide indeksering word met hierdie probleem gekonfronteer.

Cleveland & Cleveland (1990:224) wys daarop dat ten spyte van die vooruitgang op tegnologiese gebied, die mens nog nie rekenaarprogramme kan skryf wat die rekenaar in staat stel om taal op dieselfde wyse as die mens te interpreteer nie:

"A computing machine is an incredible device, but it is a poor substitute for the human brain. A computer's power lies in processing mundane trivialities (man solves a complex problem by dividing it up into a long series of mundane steps, and then the computer takes over). There is no record of any computing machine that has made a valid value judgment without human direction, and such value judgments play a major role in the creation of quality indexes and abstracts".

HOOFSTUK 3

DIE KOGNITIEWE PROSES TYDENS MENSLIKE INLIGTINGSVERWERKING EN DIE ROL DAARVAN BY INDEKSERING

3.1 Inleiding

In die vorige hoofstuk is aangedui dat die proses van onderwerpsanalise en die keuse van konsepte wat in toegekende indeksering plaasvind, hoofsaaklik 'n verstandspesies is. Hierdie hoofstuk handel oor die kognitiewe proses wat in die menslike brein plaasvind wanneer 'n dokument gelees word en inligting in die geheue verwerk en geïnterpreteer word. Verder sal aangetoon word hoe die redenasieproses verloop tydens die identifisering en keuse van konsepte wat die onderwerpsinhoud van die dokument weergee. Om hierdie proses van menslike inligtingsverwerking beter toe te lig, is dit vooraf nodig om die samestelling van die menslike brein en die natuurlike proses wat daarin tydens inligtingsverwerking plaasvind, oorsigtelik te bespreek.

3.2 Die kognitiewe proses

3.2.1 Fisiologiese samestelling van die menslike brein

Volgens Louw (1986:53) is die brein 'n komplekse struktuur en kan dit as die ingewikkeldste orgaan van die menslike liggaam beskou word. Hy sê dat die voorbeeld waar die brein met 'n rekenaar vergelyk word, baie misleidend is, aangesien die brein meer kompleks is en ook 'n veel groter kapasiteit het as enige rekenaar wat tot op hede ontwikkel is.

Vir die doel van hierdie bespreking is dit egter net nodig om kennis daarvan te dra dat die brein uit drie dele bestaan, naamlik 'n voor-, middel- en agterbrein. Die denkproses wat by indeksering plaasvind, vind in die voorbrein plaas, terwyl die proses wat plaasvind om woorde in taal uit te druk, in die agterbrein plaasvind. Hierdie drie dele van die brein werk interafhanklik saam met die limbiese sisteem wat by die geheue betrokke is

(Louw, 1986:90-91). In 3.3 sal aangetoon word hoe die menslike geheue in die berging en herroep van inligting funksioneer.

3.2.2 Natuurlike funksionering van die menslike geheue

In die natuurlike funksionering van die menslike brein, word drie dimensies van intellektuele funksionering onderskei naamlik: handelinge of aksies (die wyse waarop die intellek funksioneer), inhoud (die materiaal of inligting ten opsigte waarvan die intellek funksioneer) en produkte (die resultate van die intellektuele proses) (Louw, 1986:144-145).

Volgens Klix en Hoffman (1980:11-12) bestaan die menslike geheue uit die beskikbare data, feite en voorstellings wat deur die individu versamel is. Hierdie versameling van data verander voortdurend namate die individu met die omgewing in aanraking kom. Dit is noodsaaklik om in gedagte te hou dat die fisiese samestelling van die menslike brein die geheue omvat en dat die intellek altyd by die funksionering van die brein teenwoordig is. Enige denkproses word deur die intellek van die individu beïnvloed.

Die inligting wat in die menslike brein geberg word, het 'n drieledige oorsprong, naamlik:

- die biologiese ontwikkeling van die mens;
- die geskiedenis van die mensdom en gemeenskap;
- persoonlike ervaringe van die individu.

Elke individu se ervaring word deur middel van sintuiglike waarneming en kontak met sy omgewing verkry. Hierdie interaksie met die omgewing gee aan die individu kennis wat in die menslike geheue geberg word. Alle besluite word vanuit hierdie kennis wat die individu in sy geheue berg, geneem. Alle interne kognitiewe prosesse wat plaasvind, is gesetel tussen die bewuswording van inligting (byvoorbeeld die onderwerpsinhoud van die dokument) en aksie (byvoorbeeld die ontleding van die onderwerpsinhoud)

wat plaasvind. Hierdie twee prosedures vorm die elemente van die geheue (Klix en Hoffman, 1980:2).

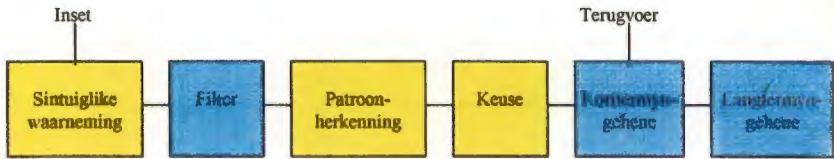
In vergelyking met die menslike geheue, beskik 'n rekenaar oor die moontlikheid om eenvoudige besluite te neem wat op syferkombinerings, karakterherkenning en die rondskuif van karakters gebaseer is. Alle "redenasies" in die rekenaar berus daarop dat twee syfers (0 en 1) relatiewe waardes het en alle data word hiervolgens voorgestel. Daar is dus geen kognitiewe proses teenwoordig nie. Uit bogenoemde is dit duidelik dat die meganiese funksionering van 'n rekenaar geensins met die interne kognitiewe prosesse wat in die menslike brein plaasvind, vergelyk kan word nie.

Vervolgens sal aangetoon word hoe die mens inligting wat uit eenhede bestaan, en nie uit syfers soos by die rekenaar nie, in kennis verwerk.

3.3 Menslike inligtingsverwerking

Die denkprosesse wat tydens indeksering plaasvind, is van menslike inligtingsverwerking afhanklik. In Figuur 1 word die proses van menslike inligtingsverwerking voorgestel.

Menslike inligtingsverwerking is die proses waartydens 'n persoon inligting verstandelik tussen 'n stimulus en 'n respons oordra. 'n Stimulus is 'n gewaarwording en 'n respons is die reaksie daarop (Farrow, 1994:155). Gouws *et al.* (1979:131) brei verder hierop uit deur te sê dat inligtingsverwerking die proses is waartydens enige inligting wat alreeds in die geheue aanwesig is, verander word of met reeds bekende inligting in konteks geplaas word om nuwe inligting te skep. Hierdie proses van inligtingsverwerking is 'n onbewustelike, aktiewe proses. Inligting word waargeneem, onthou en geberg om later weer gebruik te word. Hierdie denkhandeling wat die mens uitvoer om van inligting bewus te word, dit te onthou en om daarvoor te redeneer en tot volle begrip daarvan te kom, vind alles in die geheue van die mens plaas (Glass *et al.*, 1979:2).



Figuur 1. Fases van 'n inligtingsverwerkingsmodel (Read, 1992:5).

Interne en eksterne inligting word vanuit die omgewing deur middel van sintuie in die sensoriese register geberg. Nadat hierdie sintuiglike waarneming plaasgevind het, word die inligting deur middel van 'n filter herken. 'n Filter is 'n beheermeganisme wat die brein in staat stel om slegs bruikbare inligting as deel van 'n patroon te herken en te selekteer vir verdere verwerking (Farrow, 1994:156). Kognisie begin plaasvind, naamlik die bewuswording, herkenning en begrip van inligting deurdat voorkennis uit die langtermyngeheue na die korttermyngeheue oorgeplaas word. Hierdie denkhanelinge veroorsaak dat daar 'n verandering in die semantiese geheue begin plaasvind. Die semantiese geheue is opgebou uit 'n netwerk van knooppunte met verbindinge. Die netwerk bevat sekere katalogiserende prosedures wat aan 'n voorwerp of aksie sekere eienskappe toedig. Hierdie knooppunte verander gedurig. Die gevolg is dat die semantiese deel van die geheue gedurig veranderinge ondergaan soos wat nuwe toevoegings en integrerings plaasvind. Hierdie inligting wat in die korttermyngeheue geberg word, word óf in die langtermyngeheue geberg, óf die terugvoer kan in een of ander vorm van gedrag waargeneem word (De Wet *et al.*, 1981:114).

Die wyse waarop inligting verwerk, sinvol georganiseer en geberg word, is van groot belang aangesien dit bepalend vir die take is waarvoor die inligting in 'n latere stadium herroep en gebruik kan word. Inligtingsverwerking is 'n proses van aanhoudende inkomende inligting wat deur die geheue van 'n individu vloei. As die denkhanelinge afgehandel is, word die kognitiewe

voorkennis van die individu verander, aangesien dit met bestaande en nuwe kennis geïntegreer word en dan as nuwe bykomende kennis geberg word. Tydens inligtingsverwerking word die inkomende inligting op wisselende vlakke verwerk (De Wet *et al.*, 1981:114-115).

Dié vlakke van inligtingsverwerking wissel van oppervlakkige verwerking tot verwerking waar daar betekenis aan inkomende inligting gegee word (Andre, 1979:280-281). Volgens Craik (1979:79) en Craik & Lockhart (1972:675) het die eerste vlakke van inligtingsverwerking te doen met fisiese en sensoriese kenmerke soos 'n lyn, vlak of hoek, terwyl die dieper vlakke te doen het met 'n vergelyking tussen reeds inkomende inligting en bestaande kennis in die geheue van die individu. Hoe dieper die vlak van inligtingsverwerking is, hoe intenser onthou die individu en kan hierdie inligting later vir uitbreiding van die kognitiewe struktuur verwerk word.

Die intensiteit van inligtingsverwerking kan wissel van outomaties tot waar die aandag van die individu van die uiterste belang is. Dit word gekontroleerde inligtingsverwerking genoem word (Schneider & Shiffrin, 1977:15). De Wet *et al.* (1981:117) wys daarop dat die individu se aandag belangrik vir inligtingsverwerking is, aangesien alle inkomende inligting gesorteer en verminder kan word deur op die belangrikste aspekte te konsentreer. Die skrywers gaan ook van die standpunt uit dat as 'n individu 'n begrip van die inligting het, hy die inhoud in sy eie woorde kan weergee en opsom en dit in nuwe omstandighede kan toepas. Rowland en McGuire (1971:8) wys daarop dat die denkhandelinge wat plaasvind, inligting na 'n ander tipe vorm transformeer. De Wet *et al.* (1981:117-118) noem die denkhandeling wat plaasvind "kodering". Dit beteken dat die individu nuwe afleidings maak uit inligting wat reeds in die langtermyngeheue teenwoordig is.

Al hierdie prosesse van inligtingsverwerking vind tydens toegekende indeksering in die indekseerder se brein plaas. In vergelyking hiermee is die hele proses van afgeleide indeksering wat hoofsaaklik rekenaarmatig gedoen word, slegs op die basiese werking van 'n rekenaar gebaseer, soos wat reeds in 3.2.1 genoem is. 'n Rekenaar is nie 'n goeie plaasvervanger vir die menslike brein nie, en volgens Cleveland & Cleveland (1990:224) is daar nog

geen program geskryf wat menslike taal op dieselfde wyse as die menslike brein kan interpreteer nie.

Tydens die indekseerproses met toegekende terme is 'n indekseerder in staat om te besluit wat die "aboutness" of die *waaroor* van 'n dokument is deur die inhoud van 'n dokument op te som of kernagtig weer te gee. Hierdie proses van opsomming vind plaas deur een of ander vae proses van "vermindering" of semantiese kondensering. Die opsomming van die teksinhoud kan gesien word as 'n proses van identifisering en keuse van die onderwerpe wat 'n aanduiding is waaroor die dokument handel (Todd,1992:104).

Tydens onderwerpsanalise vind 'n bepaalde proses tydens die begryp van die onderwerpsinhoud en die identifisering en keuse van konsepte plaas. Om hierdie proses verder te verduidelik, is dit nodig om kortliks te beskryf hoe die mens taal interpreteer. Dit word gedoen volgens die teorie wat twee wetenskaplikes, naamlik T.A. van Dijk ('n taalkundige van die Universiteit van Amsterdam) en W. Kintsch ('n psigoloog van die Universiteit van Colorado), geformuleer het.

In enige teoretiese studie van die menslike intellektuele proses is 'n studie van taal nodig, want "human beings cannot communicate by direct thought transference and need some intermediate medium" (Farradane, 1980:76). Om die rede word die indekseerproses vervolgens aan die hand van die teorie van Van Dijk en Kintsch bespreek. Hierdie teorie is gebaseer op "discourse comprehension" wat met *diskoersbegrip* vertaal kan word.

3.4 Die Van Dijk-Kintsch teorie oor die interpretasie van taal

Hierdie teorie gee 'n beskrywing van wat in die menslike geheue tydens die lees en verstaan van teks plaasvind. Die onderwerp van 'n dokument word vasgestel deur die navolging van verstandelike reëls wat onbewustelik in die menslike geheue plaasvind. Dié reëls is verteenwoordigend van die mens/indekseerder se kennis oor die taal wat hom in staat stel om 'n taal aan te leer en woorde te herken deur die lees van sintakties gevormde sinne. Hierdie reëls stel die indekseerder in staat om ook die "aboutness" of die

waaroor van 'n dokument te identifiseer. Dit vind plaas deur konsepte hiërargies in die geheue te berg om sodoende 'n kernagtige opsomming van die teks op 'n diep vlak van inligtingsverwerking te maak (Beghtol (1986:90,92; Todd, 1992:104). Van Dijk en Kintsch (1983:10) stel dit soos volg:

"We go from the understanding of words, to the understanding of clauses in which these words have various functions, and then to complex sentences, sequences of sentences, and overall textual structures."

Hulle noem hierdie kognitiewe handeling om die teks kernagtig in die geheue saam te vat, makroreëls. Hierdie reëls stel die indekseerder in staat om aan 'n onderwerp 'n hoë posisie tydens die kognitiewe proses te gee sodat dit in 'n latere stadium versamel kan word. Die indekseerder is reeds in hierdie stadium in staat om die onderwerp van die dokument in een enkele konsep uit te druk (Frohmann, 1990:83).

Van Dijk & Kintsch (1983:15-16) is verder van mening dat hierdie reëls by 'n spesifieke situasie aangepas kan word en dat hulle die geheue in die identifisering van konsepte ondersteun. Die indekseerder hoef nie tot aan die einde van 'n paragraaf, hoofstuk of dokument te wag voordat hy/sy in staat sal wees om die essensie van die teks weer te gee nie. Hierdie makroreëls impliseer dat die leser uit die minimum teksinligting kan vasstel wat die onderwerp van 'n dokument is. Hierdie kennis word deur verskillende tipes inligting ondersteun, byvoorbeeld die titel, woorde wat die onderwerp omskryf, die eerste sin van 'n paragraaf, kennis oor moontlike voorafkennis en inligting in die konteks.

Die totstandkoming van teksbegrip is van twee tipes verwerking afhanklik. 'n Eenvoudige beskrywing is dat dit van bo na onder en van onder na bo plaasvind. Die bo-na-onder-verwerking is inligting wat nie in die teks verteenwoordig word nie, maar dit vorm deel van die outeur se kennis en begrip van die wêreld. In dié situasie hoop die outeur dat sy lesers óók oor die kennis en begrip sal beskik. Hierdie tipe verwerking is konseptueel omdat dit 'n uitdrukking van konsepte is wat alreeds in die geheue beskikbaar is. Inligting wat van onder na bo verwerk word, is perseptueel. Dit beteken dat die leser dit deur middel van waarneming in die korttermyngeheue

opneem (Farrow, 1991:151). Kennis van hierdie proses wat in die geheue plaasvind, is belangrik, want dit vorm 'n onafskeidbare deel van die indekseerproses waartydens die indekseerder die indekseertaal konstrueer om die inhoud van 'n dokument weer te gee, indeksinskrywings te formuleer en te korrigeer en 'n indeks fisies op te stel.

Wanneer die indekseerproses aan die hand van 'n kognitiewe model beskryf word, is dit van belang om te onthou dat daar twee onderskeibare dele is: die indekseerder se opsomming van die teks en die verteenwoordiging van hierdie opsomming deur middel van indekseerterme (Farrow,1991:151). In die proses is daar vier veranderlikes wat 'n invloed op die indekseerder se interpretasie van die dokumentinhoud kan hê:

- Indekseerders werk meestal onder druk wat beteken dat die teks oorsigtelik bestudeer word en nie volledig teen 'n normale spoed gelees word nie.
- Indekseerders bestudeer die teks uitsluitlik met oog op indeksering. Na voltooiing van die taak het die indekseerder geen verdere belang by die teks nie.
- Nadat 'n begrip van die dokumentinhoud verkry is, is die skep van 'n stel indekseerinskrywings die volgende stap. Daar vind dus onmiddellik 'n aktiwiteit plaas.
- Indekseerders werk in 'n beperkte veld wat dokumentsoort en onderwerpsveld betref. Dit bring 'n herhalings-element mee wat daartoe bydra dat teksbegrip outomaties plaasvind, wat nie die geval met gewone lees is nie.

Farrow (1994:159-164) gebruik Kintsch en Van Dijk se teorie om aan te dui hoe 'n indekseerder 'n begrip van die inhoud van 'n dokument vorm. Volgens hom kan hierdie teorie as uitgangspunt gebruik word om die proses van indeksering te beskryf. 'n Teks bestaan, soos reeds gemeld, uit die opeenvolging van woordvoorstellings wat dieselfde idee weergee. 'n Voorstelling is die kleinste begripseenheid wat ontleed kan word. 'n Leser kan tydens die lees van 'n paragraaf besluit wat die tema of onderwerp is. Volgens hierdie teorie sal die indekseerder in staat wees om dit in een

aanvaarbare sin op te som sonder om intensief daarvoor te gaan sit en nadink. As die indekseerder byvoorbeeld die volgende paragraaf uit Louw (1986:395) lees: "Om 'n beter begrip van slaap as droom te verkry, word vervolgens (afsonderlik) op die belangrikste aspekte van die twee bewussyntoestande gelet", weet die indekseerder onmiddellik dat dit oor slaap en droom gaan en dat die twee prosesse verskil. Die indekseerder identifiseer dus die onderwerp, som dit in een sin op sonder om na te dink en verbind dan hierdie sin met ander sinne wat in die teks voorkom. Vir die indekseerproses het dit tot gevolg dat 'n indekseerder onmiddelik weet waaroor 'n dokument handel, al konsentreer hy slegs op 'n paar sinne in die dokument.

Die voorstellings wat die indekseerder oor slaap en droom maak, vind hiërargies in die geheue plaas en hy kan hulle op dieselfde manier in die teks self begin interpreteer. Die struktuur wat in die teks deur die oorspronklike outeur gevorm word, hou gewoonlik met die titel van die dokument verband.

Wanneer 'n teks bestudeer word, verbind die leser hierdie voorstelling met mekaar en verbind hy hulle direk met inligting wat uit sy geheue afkomstig is. Die inligting het 'n drieledige oorsprong :

- Indien die indekseerder nie in staat is om hierdie voorstellings met mekaar te verbind nie, word inligting vanuit die langtermyngeheue opgeroep.
- Daar vind onbewustelik 'n proses plaas waarin die inligting in die teks só verwerk word dat sekere belangrike voorstellings behou word en so ver as wat die indekseerder lees aan die volgende voorstellings verbind word.
- Hierdie proses van die verwerking van voorstellings vind deurlopend plaas totdat die konsep met konsepte in die indekseertaal verbind word.

Volgens Louw (1986:265) beskik die mens oor 'n konseptuele stelsel vir probleemoplossing. Hierdie stelsel word gevorm uit die interaksie met kultuur, aangeleer deur opvoeding en elke individu se stelsel is op sy eie ervaring en oordeel gebaseer. Die indekseerder gebruik hierdie konseptuele

stelsel tydens die ontleding van 'n dokument se inhoud om te bepaal wat met die onderwerp verband hou en wat nie.

Samevatting

Skrywers soos Farradane (1980:75) en Farrow (1991:164) is van mening dat meer navorsing gedoen behoort te word oor die denkproses van die mens om 'n duideliker begrip oor die skep en ontvangs van inligting te hê. Kennis hieroor sal navorsers in staat stel om die hele proses van inligtingsverwerking en -herwinning effektiewer vir beter resultate te kontroleer. Farradane wys ook daarop dat die inligtingswetenskap 'n kognitiewe wetenskap is wat hom met die denkproses wat in die mens plaasvind, bemoei. Kennis wat uit die bestudering van die denkproses verkry word, kan daartoe bydra dat inligtingsverwerkingsmodelle vir rekenaars ontwikkel kan word. 'n Belangrike bydrae sal gelewer kan word deur verskillende teoretiese uitgangspunte te gebruik waarvolgens die denkproses wat nie menslik waargeneem kan word nie, te beskryf en toe te pas deur dit konkreet in 'n rekenaar na te boots (De Wet *et al.*, 1981:118).

Navorsing word tans meer op die tweede stap van die indekseerproses (naamlik die identifisering van konsepte) gerig, terwyl die eerste stap (naamlik die analisering van die dokument en die vasstelling van die onderwerpsinhoud) nie soveel aandag geniet nie. Volgens Frohmann (1990:83) kom dit voor asof navorsers dit vermy om oor die eerste stap te skryf. Ook ander outeurs steun hierdie standpunt. Todd (1992:102) sê:

"The current literature of indexing does not address how the indexer is expected to be able to undertake the process of subject analysis. There is need to understand the rules, procedures that govern the mental activities of subject analysis . . . How are indexers able to write down the subject of documents? What internal guidance systems determine their actions? What is the cognitive process of identifying the intrinsic aboutness of documents, that is the cognitive process by which an indexer describes the intrinsic subject or aboutness of a document?"

Farradane (1979:5) se kommentaar hierop is: "The indexer's 'output' is a stylised form of 'information' whose relationship to the original document needs more study than it usually receives. "

Frohmann (1990:82) is ook van mening dat 'n intellektuele proses onderskei kan word waarin konsepte in die geheue gevorm word om 'n dokument deur middel van 'n indekseerterm te verteenwoordig. Die eiesoortige aard van hierdie proses is egter nog nie beskryf nie. Indekseerders is geneig om te dink dat as indekseerreëls verstaan en toegepas word, die indekseerterme 'n logiese deel van die eindproduk is. Frohmann is ook van mening dat as daar meer klem daarop gelê word dat hierdie proses verstaan moet word, 'n belangrike bydrae tot die inligtingswetenskap gelewer sal word en baie van die kwelvrae rondom indeksering opgelos kan word.

3.5 Stappe in die indekseerproses

Vervolgens sal die stappe in die indekseerproses soos dit kortliks in 2.2.2 genoem is, in meer besonderhede bespreek word om aan te toon hoe die kognitiewe proses daar plaasvind. Die oorkoepelende stappe (daar is ook substappe, soos later sal blyk) is die volgende:

- analisering van die dokument en vaststelling van die onderwerpsinhoud;
- identifisering en keuse van konsepte in die dokument;
- oordrag van hierdie konsepte na die terme van die indekseertaal.

Soos reeds aangetoon, oorvleuel die eerste twee stappe meestal en staan dit bekend as die proses van konseptuele analise. Volgens Lancaster (1991:8) behoort die indekseerder homself die volgende vrae tydens konseptuele analise af te vra:

- Waaroor handel die dokument?
- Waarom is die dokument tot 'n spesifieke versameling gevoeg?
- In watter aspek van die onderwerp sal gebruikers geïnteresseerd wees?

Die doel van onderwerpsanalise tydens indeksering is die formulering van indekseerterme wat in die derde fase volvoer word. In die geval van toegekende indeksering moet dit volgens die indekseerstelsel se voorskrifte gekontroleer word. Hierdie hele proses sal teen die agtergrond van menslike inligtingsverwerking bestudeer word, soos bespreek is in 3.3 (Farrow, 1994: 156-158).

Dit is belangrik om in gedagte te hou dat voordat daar na die eerste stap van indeksering oorgegaan word, die indekseerder alreeds 'n keuse maak deur te besluit of dit die moeite werd is om die dokument te indekseer. Dit word gedoen deur dit oorsigtelik te bestudeer. Indekseerbesluite word gebaseer op die doelstellings en beleid van die instansie wat die indeksering onderneem. Die beleid van 'n groot indekseerinstansie is gewoonlik toegespits op 'n uiteenlopende gebruikersgroep en indeksering vind dan gewoonlik oor 'n wye veld van onderwerpe plaas. Die teenoorgestelde sal binne 'n kleiner organisasie geld. Indeksering word gedoen volgens die spesifieke behoeftes van die gebruikers met die gevolg dat daar onderwerpsgespesialiseerd te werk gegaan word. Die behoeftes van gebruikers bepaal die aard van die indekseerproses wat diepte en spesifiekheid betref. Met inagneming van al hierdie faktore vind die proses van indeksering plaas.

Vervolgens word elke stap asook substappe bespreek.

3.5.1 Analisering van die dokument en die vasstelling van die onderwerpsinhoud

Sodra die indekseerder die dokument begin deurlees, begin sekere gedagtes in sy brein oor die inhoud posvat. Farrow (1994:158-159) wys daarop dat dit nie altyd moontlik is om 'n dokument volledig deur te lees nie. Die *UNISIST Indexing Principles* gee leiding in hierdie geval deur die belangrike dele van die teks wat volledig bestudeer behoort te word, te noem. Dit sluit die volgende in: die titel, ekserp (indien verskaf), inhoudsopgawe, inleiding, inleidende en slotparagraaf, illustrasies, onderstreepte woorde of woorde waar die lettervorm verskil. Die outeur se doelstellings en die mate waarin hierdie doelstellings bereik is, kom gewoonlik in die inleidende dele asook in

die slot na vore. Die indekseerder se begrip word nie net deur die strukturele eienskappe van die dokument (naamlik inleiding, gevolgtrekking, ensovoorts) verdiep nie, maar die frekwensie van woorde in die teks en die semantiese verwantskappe tussen woorde speel ook 'n belangrike rol. As 'n indekseerder byvoorbeeld 'n woord soos "carpets" teëkom, verbind hy outomaties woorde soos "floorcoverings", "rugs" daaraan en besluit watter woord die beste by sy eie semantiese netwerk van konsepte inpas (Farrow, 1991:154).

Volgens Cleveland & Cleveland (1990:105) is dit nodig dat die indekseerder moet besluit watter aspekte van 'n onderwerp beklemtoon moet word en op watter aspekte minder klem gelê moet word. Hierdie besluite sal afhang van die behoeftes van gebruikers en word alreeds tydens hierdie stadium van die analisering en vasstelling van die onderwerpsinhoud geneem.

Nadat die indekseerder bogenoemde besluite geneem het, begin hy konsepte neerskryf wat óf direk uit die teks voorkom óf in sy eie woordeskat teenwoordig is óf die resultaat van 'n kombinasie van albei is.

3.5.2 Begrip van die dokumentinhoud

Farradane (1980:75-76) dui aan dat die totstandkoming van gedagtes tydens die analise van 'n dokument 'n intellektuele proses is wat in die geheue plaasvind. 'n Indekseerder verkry reeds in hierdie stap 'n idee van die onderwerpsinhoud deur dele van die teks te ontleed soos in 3.4 bespreek is. Hierdie gedagtes neem vorm aan voordat dit in taal uitgedruk word en is die konsepte wat geïdentifiseer word. Die indekseerder gebruik taal om hierdie gedagtes uit te spreek. Hierdie stap vind instinktief plaas, al word hulle nie fisies in hoorbare klanke uitgedruk nie. Alle individuele konsepte wat die indekseerder identifiseer, tree as boustene van gedagtes op en individuele woorde word gebruik om aan hierdie konsepte benaminge toe te ken. Hierdie konsepte word volgens die funksionering van die geheue aan mekaar verbind. Die denkhandeling wat plaasvind om hierdie konsepte met mekaar in verband te bring, moet nie met die grammatika en die sintaksis (of die reëls) van taal verwar word nie, aangesien dit van taal tot taal verskil. Sommige van hierdie taalreëls beheer die oordrag van gedagtes in taal.

Volgens die *UNISIST Indexing Principles* is insig in die dokumentinhoud van kardinale belang voordat enige konsep geïdentifiseer kan word. Volle insig is in 'n sekere mate afhanklik van die uiterlike voorkoms van die dokument. Dokumente kan in geskrewe of ongeskrewe vorm wees. Meestal is dit moeiliker om tot volle insig by ongeskrewe dokumente te kom (byvoorbeeld films, klankopnames), omdat hulle gewoonlik moeiliker as geskrewe dokumente hanteer en addisionele apparaat vereis. Geskrewe dokumente kan intensiewer bestudeer word. Die uiterlike formaat waarin 'n dokument voorkom, is egter nie vir die doel van hierdie studie belangrik nie, aangesien daar slegs op die indekseerproses en die stappe wat in die proses gevolg word, gekonsentreer word.

3.5.3 Identifisering van konsepte

Nadat 'n dokument bestudeer en ontleed is en die indekseerder besluit het waaroor dit handel, moet konsepte geïdentifiseer word en in woorde uitgedruk en gelys word. Volgens Cleveland & Cleveland (1990:106-107) is die doelstelling van die indekseerder in hierdie stadium om in sy geheue 'n oorsig te vorm oor wat die outeur probeer sê. Probleme is dat die outeur nie fisies teenwoordig is nie en die indekseerder van die swart gedrukte simbole wat op die papier voorkom, gebruik moet maak. Taal kan ook verwarrend en metafories wees. In hierdie stadium moet die indekseerder die toepaslikste besluite oor die onderwerp begin neem. Die onderwerpsfrase "The mosquitos attacked with the ferocity of a tiger" handel byvoorbeeld oor muskiete en nie oor tiere nie - *tiere* sal dus nie die korrekte term wees om hier te gebruik nie. Die indekseerder begin onbewustelik intellektuele besluite te neem wat vir die indeksering van die dokument bepalend sal wees.

3.5.4 Die keuse van konsepte

Dit is noodsaaklik dat 'n indekseerder ter wille van konsekwentheid 'n logiese benadering moet volg om konsepte wat die onderwerpsinhoud die beste uitdruk, te kies. Cleveland & Cleveland (1990:107) dui aan dat die

fundamentele probleem hier is dat die konsep van 'n onderwerp misleidend kan wees. Skrywers heg soms 'n eie konnotasie aan 'n sekere onderwerp en dit beïnvloed die wyse waarop hulle 'n onderwerp omskryf.

Die keuse van konsepte word in hierdie stadium weergegee in die indekseerder se eie woorde of woorde wat uit die dokument ontleen word. Hierdie lys konsepte moet so volledig moontlik wees, maar ook so presies moontlik. Die keuse van konsepte kan ook aan die hand van 'n skema of hiërargie van verwante faktore wat die onderwerpsveld dek, gedoen word. Konsepte word aan die hand van onder meer gebeure, prosesse, toerusting en navorsing geïdentifiseer (UNESCO, 1975:5).

Om te besluit watter konsepte belangrik is, kan 'n eenvoudige reël in gedagte gehou word, naamlik dat hoe meer 'n idee in die teks teenwoordig is, hoe groter is sy belangrikheid (Cleveland & Cleveland, 1990:107). Die meeste tegnieke wat in afgeleide indeksering gebruik word, is gebaseer op die frekwensie van woorde in die teks. Dit is as gevolg hiervan dat eerder hoofidees as minder belangrike idees gekies word. Verder moet die styl en grammatiese struktuur ook in gedagte gehou word. In die sin "The mosquito attacked with ferocity" is die woord "mosquito" belangrik, maar in die sin "The queen looked at me with her mosquito eyes" is "mosquito" van minder belang om die hoofidee vas te stel.

Farrow (1994:162) beklemtoon dat indeksering op 'n kognitiewe grondslag plaasvind:

"These factors owe an obvious debt to the standard citation order of bibliographic classification theory, and are evident of the conceptual identity of indexing and classification."

Vervolgens sal aangetoon word hoe die konsepte wat geïdentifiseer is, na die terme van die indekseertaal oorgedra word.

3.5.5 Oordrag van konsepte na die indekseertaal

Die laaste stap in die indekseerproses is die oordrag van die konsepte wat gekies is, na die terme van die indekseertaal. Die konsepte moet oorgedra

word na 'n lys bruikbare indekseerterme wat verteenwoordigend van die dokument se inhoud sal wees. In toegekende indeksering vind hierdie oordrag plaas aan die hand van 'n gestandaardiseerde lys van terme soos wat reeds volledig in 2.4 bespreek is. Volgens Cleveland & Cleveland (1990:107) is indeksering nie net 'n blote roetinehandeling nie, maar is dit 'n kombinasie van formele reëls, logiese redenering en talent wat deurgaans toegepas word:

- Die indekseerder kom tot die besef dat gebruikers moontlik onder ander terme sal soek of 'n ander konnotasie aan 'n woord as hy sal heg.
- Die indekseerder moet van hulpmiddele soos 'n tesourus of saakhooflyfs gebruik maak met die oog op gestandaardiseerde terme (soos in 2.4.2.1 verduidelik is).
- As 'n bepaalde term nie in 'n lys voorkom nie, moet aan 'n sinoniem gedink word. Kruisverwysings moet gemaak word om hulp en leiding aan gebruikers te verskaf en om verwarring uit te skakel wat variante spelwyses en plekname betref.

Samevatting

Anderson (1985:287-323) vat die proses wat in die menslike geheue plaasvind en die parallelle indekseerproses raak deur die beeld van 'n biblioteek te gebruik. Volgens hom is die mees gekompliseerde en gesofistikeerde indekseerstelsel dié wat deur die mens in sy eie geheue geskep word deur sy eie "breindokumente" te indekseer en hierdie dokumente deur onderlinge breinsenuwees aan mekaar te verbind.

Die mens se geheue is terselfdertyd 'n bergings- en kommunikasiestelsel. Deur inligting te indekseer, dit wil sê kernagtig deur konsepte te verteenwoordig, word die bergings- en kommunikasiedeel van die geheue in 'n inligtingsherwinningstelsel verander. Hierdie verandering vind plaas deur die ontvangs van simbole deur persepsie en die verbinding van hierdie konsepte met mekaar. Denkhandelinge oefen beheer uit deur te besluit watter

van hierdie simbole belangrik is en watter nie. Die simbole wat gekies word, word dan in die "breinindeks" van die geheue geberg.

Die brein funksioneer net soos wat 'n goed beplande biblioteek bestuur word. Dit is in staat om te onderskei tussen bronne wat permanent tot die versameling gevoeg moet word en dié wat nie permanent opgeneem moet word nie. Die geheue kies dus "n gebruikversameling" deur 'n klein deeltjie van hierdie bronne vir verwerking na die langtermyngeheue toe te stuur. Hierdie konsepte in die "gebruikversameling" word dan in die indekseerder se geheue geberg as deel van 'n inligtingsherwinningstelsel. Net soos wat 'n goed beplande biblioteek alle dokumente wat nie gebruik word nie, sal afskryf, sal die menslike geheue alle onbelangrike inligting vergeet. Alhoewel hierdie inligting op dieselfde tydstip as die belangrike inligting in die langtermyngeheue geïndekseer is, is dit minder belangrik geag toe dit ontvang is en gevolglik nie so intensief geïndekseer nie.

Die menslike geheue is net soos 'n biblioteekversameling of 'n voltekstdatabasis waar dokumente volledig in die indekseerstelsel ingevoeg word, maar nie deur kernagtige surrogate verteenwoordig word nie. Die geheue is nie in staat om die oorspronklike teks wat deur die geheue verwerk word, te berg nie. Die geheue tree selektief op deur bruikbare konsepte uit die massas inligting wat ontvang word, te kies. Hierdie bruikbare konsepte vorm die "nuwe" dokumente wat geskep word en in die geheue geïntegreer word. Dit is die volledige teks van hierdie nuwe dokumente wat deur die geheue geberg en geïndekseer word.

3.6. Die soekproses

In 2.1 is gesien dat een van die doelwitte van indeksering is om soekpunte in die katalogus of indeks te bied wat die gebruiker na die inligting moet lei. Hierin speel die indekseertaal 'n deurslaggewende rol. Om te kan bepaal of die indeksering geslaag is en sy doelwit bereik en of die indekseertaal daarin slaag om relevante dokumente op te spoor, is dit nodig om ook kortliks te kyk wat tydens die soekproses plaasvind. Dit is opmerklik dat die soekproses in

baie opsigte met die proses van toegekende indeksering ooreenstem, omdat daar ook 'n kognitiewe proses teenwoordig is, soos wat aangetoon sal word.

As inligtingsherwinning plaasvind, moet dieselfde woorde wat deur die indekseerder gebruik is om die inhoud van 'n dokument te beskryf, tydens die navraagproses gebruik word om inligting te herwin. Om die rede speel die gebruiker so 'n belangrike rol - die stelsel se sukses word grootliks gemeet aan die mate waarin dit aan gebruikers se behoeftes voldoen.

Vickery & Vickery (1987:127-129) beskryf die soekproses as die ontstaan van 'n inligtingsbehoefte in die geheue van die gebruiker. Dit word vergestalt as 'n navraag in natuurlike taal aan homself of 'n bibliotekaris. Hierdie gestelde of gestruktureerde vraag verskil egter van 'n vraag wat normaalweg in geselstrant aan individue gestel word, omdat dit uit 'n inligtingsontsluiting-herwinningstelsel (IOHS) beantwoord moet word. Dit verskil op grond van die sintaktiese struktuur en gestandaardiseerde terminologie word gebruik om 'n gestelde navraag te analiseer, omdat 'n inligtingsbehoefte gewoonlik in die meeste gevalle verder uitgebrei of gewysig moet word.

Volgens Lancaster (1979:147) kan die soekproses deur die volgende veranderlikes beïnvloed word :

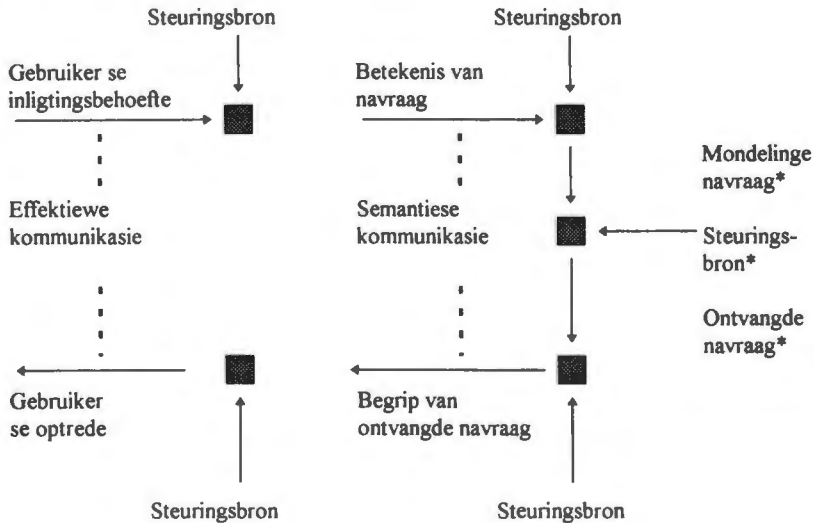
- die gebruikers se vermoë om self 'n inligtingsbehoefte in sy geheue te identifiseer en dit in woorde uit te druk;
- die verwagtinge wat die gebruiker het van moontlikhede wat die stelsel bied en die gepaardgaande ondersteuning wat deur die stelsel verskaf word.

Dieselfde kognitiewe proses wat tydens indeksering plaasvind, kan ook in die soekproses waargeneem word. Kuhlthau (1991:336-338) is van mening dat die navraagproses ook 'n studie van menslike inligtingsverwerking is, soos wat in 3.3 bespreek is. Die inligtingsbehoefte van gebruikers kan net verstaan word as hulle in konteks gebring word met die psigologiese faktore wat 'n behoefte in die eerste plek laat ontstaan. Sodoende sal insig verkry kan word in die wyse waarvolgens 'n navraag geformuleer word.

3.6.1 Stappe in die soekproses

Die soekproses bestaan uit 'n reeks opeenvolgende stappe waartydens daar kommunikasie plaasvind tussen die navraer (gebruiker) en die inligtingsontsluiting- en -herwinningstelsel (IOHS). In hierdie kommunikasieproses kan 'n tussenganger, naamlik die bibliotekaris, ook betrek word.

Harter (1986:15-16) pas die Shannon-Weaver-kommunikasiemodel op die kommunikasieproses toe om die interaksie tussen die navraer en die tussenganger te beskryf (vergelyk Figuur 2).



* Tegnieuse kommunikasie

Figuur 2. Kommunikasie tussen die gebruiker en die bibliotekaris (volgens Shannon en Weaver) (Harter, 1986:17)

Die stappe wat deur die bibliotekaris gevolg word om die inligtingsbehoefes van die gebruiker te bevredig, bepaal die geslaagdheid van die kommunikasieproses. Die kommunikasie wat plaasvind kan net geslaag wees as daar akkurate semantiese en tegniese kommunikasie is. Semantiese kommunikasie vind plaas wanneer die bibliotekaris die presiese betekenis van die geskrewe of mondelinge boodskap van die gebruiker ontvang. Tegniese kommunikasie het te doen met die tegniese aspekte van kommunikasie, soos byvoorbeeld 'n onduidelike uitspraak, onleesbare handskrif en steurende agtergrondsgeluide.

Die soekproses is net soos die indekseerproses afhanklik van stappe wat deur 'n tussenganger gedoen word om te sorg dat 'n navraag reg vertolk word en om die betekenis korrek en sinvol uit te druk. Hierdie stappe behels:

- 'n behoefte aan inligting wat op 'n navraag uitloop;
- die analisering van die navraag;
- die oordrag van die navraag na die terme van die indekseertaal;
- die soek en beantwoording van die navraag vanuit die IOHS.

Harter (1986:17) verwys na die stappe in die soekproses as take. Die eerste taak word uitgevoer wanneer die gebruiker se navraag bepaal word en die konteks daarvan verstaan word. Die bibliotekaris maak dan gebruik van interpersoonlike kommunikasievaardighede om die gebruiker op sy gemak te stel sodat die werklike omvang van sy behoefte bepaal kan word. Taylor (1968:109) identifiseer vier vlakke waarop navrae gestel kan word, naamlik:

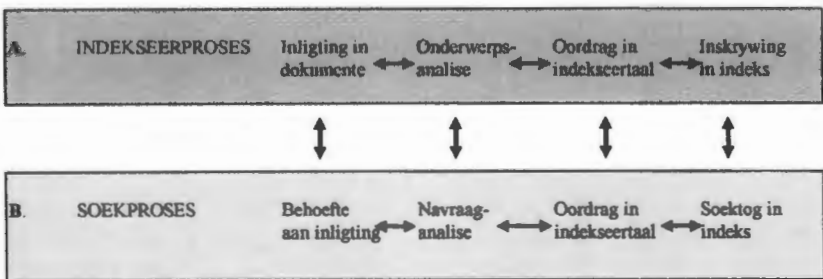
- 'n aangevoelde behoefte wat as 'n navraag uitgedruk word;
- 'n bewuste behoefte wat nog glad nie uitgedruk kan word nie en dubbelsinnig kan wees;
- 'n formele behoefte wat presies en ondubbelsinnig is;
- 'n aangepaste behoefte wat geïdentifiseer word wanneer bogenoemde drie tipes behoeftes so verwoord is dat dit deur die IOHS beantwoord kan word.

Nadat die bibliotekaris 'n begrip gevorm het van die probleem en die literatuur wat daarmee verband hou, is hy eerstens in staat om die navraag in frases wat konsepte beskryf en na simbole van konsepte om te skakel en tweedens om 'n soekstrategie te formuleer.

Die bibliotekaris as tussenganger in die kommunikasieproses moet ook oor sekere vaardighede beskik om die soekproses en die navraagonderhoud suksesvol te laat verloop. Volgens Armstrong & Large (1988:1) moet die bibliotekaris oor die vermoë beskik om interpersoonlik met gebruikers te kan kommunikeer en die vaardigheid hê om insig in die inligtingsbehoefte van 'n gebruiker te kan verkry. Verder moet die bibliotekaris ook oor die vaardigheid en kennis beskik om die terminologiese struktuur van 'n IOHS te verstaan, asook die wyse waarvolgens onderwerpe in die herwinningstelsel georden en gerangskik is.

3.6.2 Ooreenkoms tussen die indekseerproses en die soekproses

Langridge (1989:5-6) wys op parallelle stappe in die soek- en indekseerproses soos in Figuur 3 aangedui word. Rowley (1987:176) deel die opvatting van parallelisme deur die volgende stelling te maak: "Indexing and searching then, are integral one to another . . ."



Figuur 3. Ooreenkoms tussen die indekseerproses en die soekproses (Langridge, 1989:5).

Die indirekte doel van die stappe wat tydens die indekseerproses en die soekproses uitgevoer word, is om die verwoorde inhoud van die denkprosesse van individuele skrywers sistematies te onsluit sodat dit herwin kan word.

Die eerste stap tydens die indekseerproses vind plaas as die kennis wat in dokumente verteenwoordig word, ontsluit word. Die indekseerder kry tydens hierdie stadium insig in die dokument. Die soekproses word grootliks gebaseer op die onsekerheid wat by 'n gebruiker ontstaan as hy agterkom dat daar 'n gaping in sy kennisstruktuur is. Hierdie gaping veroorsaak dat hy 'n behoefte aan inligting het wat bevredig moet word uit kennis wat in dokumente verteenwoordig word.

Die tweede stap hou in dat die onderwerpsinhoud tydens die indekseerproses geanaliseer moet word. Tydens hierdie stap word die inhoud van 'n dokument ontleed en die vernaamste konsepte word geïdentifiseer. Dieselfde vind plaas tydens die soekproses as 'n gebruiker se inligtingsbehoefte geïdentifiseer word en in woorde uitgedruk word. Die ooreenkoms tussen die indekseer- en soekproses lê daarin dat alle inligting wat met 'n konsep verband hou, geïdentifiseer word sodat dit later sistematies en geordend in die vorm van 'n navraag of 'n indekseerterm weergegee kan word. Tydens albei hierdie stappe word 'n profiel opgebou. Die indekseerder bou 'n dokumentprofiel op deur verwante onderwerpe met mekaar in verband te bring, terwyl die soekprofiel weer toegespits is op 'n samevoeging van hierdie terme wat gebruik word om die onderwerpe met mekaar in verband te bring.

Die derde stap vind plaas wanneer konsepte tydens die indekseer- en soekproses na dié van die indekseertaal oorgedra kan word. Tydens hierdie stap moet die konsepte wat tydens die indekseerproses gekies is, oorgedra word as 'n lys van bruikbare terme wat verteenwoordigend van die dokument se inhoud is. Die indekseerder moet vertrou wees met die woordeskat van die indekseertaal én die beginsels waarvolgens die onderwerpe georden word as hulle in die indeks gekonstrueer word. Die navraag wat tydens die soekproses geanaliseer is, word omgeskakel na die indekseerterme van die IOHS wat gebruik word. Die gebruiker kan dus inligting herwin deur in sy

inligtingssoektog op dieselfde aspek te konsentreer as waarop die indekseerder hom toegespits het.

Samevatting

Uit die bespreking in hierdie hoofstuk blyk dit duidelik dat daar tydens die indekseer- sowel as die soekproses, baie menslike denkprosesse plaasvind: die verkryging van begrip en insig, die identifisering van konsepte en die keuse van konsepte en indekseerterme. Daar vind ook menslike kommunikasie plaas. Die afleiding kan dus gemaak word dat toegekende indekseerterme aansienlik van afgeleide terme sal verskil. Verder sal die wyse waarop 'n gebruiker die soekproses benader, baie met die wyse waarop die indekseerder die indekseerproses benader, ooreenstem.

HOOFSTUK 4

DIE KWALITEIT VAN INDEKSERING EN INDEKSEERTERME

4.1 Inleiding

Die hoofdoel van indeksering is dat 'n gebruiker in staat gestel moet word om dokumente te herwin wat oor 'n spesifieke onderwerp handel waarin die gebruiker geïnteresseerd is. Cleveland & Cleveland (1990:144) beklemtoon veral gebruikerstevredenheid as die einddoel van indeksering en wat as die oorkoepelende doel van die inligtingsproses gesien moet word. Gebruikerstevredenheid impliseer dat die inligtingsontsluiting- en -herwinningstelsel (IOHS) só moet presteer dat relevante dokumente tydens 'n soektog herwin en nierelevante dokumente terughou word. Die prestasie van 'n IOHS is dus ten nouste met gebruikerstevredenheid vervleg.

Cleveland & Cleveland (1990:144) sê verder dat al kan die toekenning van indekseerterme op verskillende wyses plaasvind (deur 'n indekseerder of rekenaar), die eindproduk in albei gevalle 'n bruikbare indeks moet wees. Presiesheid en akkuraatheid tydens die invoer stadium wanneer die dokumentinhoud beskryf word, bepaal die kwaliteit van die indeks as dit by die herwinning van relevante dokumente kom. Om die kwaliteit van die indeks vas te stel, is dit nodig dat die IOHS geëvalueer word. Die kwaliteit van indeksering kan egter volgens Lancaster (1991:77) net ex post facto geëvalueer word en daar is nie so iets soos 'n stel korrekte terme nie. Hy gaan van die standpunt uit dat daar 'n objektiewe wyse moet wees waarvolgens die kwaliteit van indeksering vasgestel kan word. Wat egter ook in gedagte gehou moet word, is dat die kwaliteit van indeksering beïnvloed word deur die tipe indekseermetode wat gebruik is.

Daar is verskillende meeteenhede waarvolgens 'n IOHS geëvalueer kan word. Sommige meeteenhede is gerig op kwantiteit en ander op kwaliteit. Vir gebruikerstevredenheid is kwantiteit sowel as kwaliteit van belang. Om werklik te bepaal of 'n gebruiker met 'n inligtingsoektog tevrede is, is moeilik en uiters subjektief. Daar is dus geen meeteenheid wat gebruikerstevredenheid kan bepaal nie. Daar kan egter uit die

evalueringsresultate afleidings gemaak word van hoe die IOHS presteer. Indien dit goed presteer, kan die gevolgtrekking gemaak word dat gebruikers tevrede behoort te wees. Verder kan ook uit die resultate by evaluering afleidings oor die kwaliteit van die indekseerterme gemaak word: as 'n soektog onder sekere terme voldoende relevante dokumente opspoor, kan gesê word dat die terme die onderwerp korrek uitdruk en dat die indeksering goed is.

Die kwaliteit van indeksering en indekseerterme beïnvloed ook die prestasie van die totale IOHS. Vervolgens sal eers na hierdie invloed gekyk word.

4.2 Faktore wat die prestasie van die stelsel beïnvloed

4.2.1 Akkuraatheid van indeksering

Die mate waarin die indekseerder daarin slaag om die onderwerpsinhoud van 'n dokument suksesvol deur middel van 'n stel terme weer te gee, bepaal die akkuraatheid van indeksering.

Lancaster (1979:196) is van mening dat die indekseerder tydens die indekseerproses beheer uitoefen oor die akkuraatheid waarvolgens terme gekies word. Cleveland & Cleveland (1990:109) en Aluri *et al.* (1989:56-58) bevestig dit deur te sê dat die indekseerder eiehandig die kwaliteit van indeksering bepaal deur die besluite wat hy neem en dat hierdie indekseerbesluite goed deurdink moet wees. Indekseerbesluite kan egter beïnvloed word deur faktore soos :

- die mate waarin die indekseerder tydens die bestudering van die dokumentinhoud konsentreer;
- die kennis wat die indekseerder oor die onderwerp of vakgebied het;
- die toegewydheid waarmee geïndekseer word;
- die akkuraatheid by die kies van relevante terme, spelling en punktuasie.

Dié faktore kan daartoe lei dat die indekseerder tydens die indekseerproses óf 'n term of terme uitlaat wat verteenwoordigend van kardinale onderwerp(e) in die dokument is, óf 'n term gebruik wat nie beskrywend van die inhoud is nie.

4.2.2 Uitputtendheid van indeksering

Uitputtendheid is 'n tweede faktor wat die prestasie van indeksering beïnvloed. Hiermee word bedoel hoe volledig 'n dokument geïndekseer word. Harter (1986:114) se definisie van die volledigheid van indeksering lui soos volg:

"Exhaustivity of indexing refers to the number of concepts treated in a document (or other entity) that are indexed by the database producer. If 'all' concepts of a document are indexed, the indexing is said to be exhaustive. If only major concepts are indexed, the indexing is less exhaustive. If only a single subject heading or classification code is assigned (to) each document, then the indexing is not exhaustive at all."

Uitputtendheid hou ook verband met die vermoë van die indekseerder om alle onderwerpe wat in 'n dokument onderskei kan word, tydens die proses van konseptuele analise volgens 'n beskrywende konsep te identifiseer en hierdie konsepte oor te dra na die terme van die indekseertaal. Die vlak van uitputtendheid waarvolgens geïndekseer word, is 'n belangrike maatstaf wat herwinning beïnvloed, alhoewel dit geensins verband hou met die wyse waarvolgens die indekseerder indeksering toepas nie. Die uitputtendheid van indeksering is grootliks afhanklik van bestuursbesluite - tydens die proses van konseptuele analise, die tipe inligtingsinstansie en die versameling dokumente. 'n Spesiale biblioteek wat byvoorbeeld op chemiese navorsing ingestel is, sal die bronne wat in sy besit is baie intensief indekseer, aangesien die gebruikers navorsers is wat in gespesialiseerde feite en inligting belang stel. In vergelyking hiermee sal 'n universiteit se natuurwetenskappe-biblioteek hulle op 'n algemener vlak indekseer omdat die gebruikers 'n heterogener groep is. Die vlak van uitputtendheid en volledigheid waarvolgens geïndekseer word, hou ook nie verband met die eienskappe van die indekseertaal nie. Dit kan egter verband hou met die tipe dokumente wat

deur die instansie geïndekseer word: interne publikasies en verslae sal uitputtender geïndekseer word.

Die volgende is ook van belang:

- Die mate van insig wat 'n indekseerder in 'n onderwerp het, kan veroorsaak dat 'n minder bekende onderwerp minder intensief ontleed word.
- Hoe meer kennis die indekseerder in 'n bepaalde kennisveld het, hoe beter sal hy aspekte uit daardie kennisveld kan identifiseer, byvoorbeeld op wetenskaplike, sosiale, tegniese en ekonomiese terrein. Hierdie kennis wat die indekseerder in sy geheue het, kan ook daartoe bydra dat inligting geïdentifiseer word wat vir verskillende gebruikersgroepe toepaslik is.

4.2.3 Eenvormigheid by die keuse van terme

Eenvormigheid impliseer dat belangrike konsepte wat in 'n dokument verteenwoordig word, so uitgedruk word dat dieselfde terme deurgaans deur verskillende persone gebruik word (Leonard, 1977:2,29-50). Terme moet dus gestandaardiseer en konsekwent gebruik word.

Volgens die *UNISIST Indexing Principles* word die keuse van konsepte beïnvloed deur :

- die mate van opleiding en ervaring van die indekseerder;
- die noukeurigheid verantwoordelikheid waarvolgens die indekseerder te werk gaan om terme in 'n gesagslys te kontroleer;
- die beleid van die instansie ten opsigte van die gebruik van indekseerhulpmiddele;
- die toepassing van voorskrifte ten opsigte van die lengte van 'n dokument, diepte van indeksering en die mate van kontrole.

In die keuse van terme kan terme te eng of te breed gekies word; verwarring ten opsigte van plekname kan ontstaan; variante spelwyses kan ignoreer word; ten slotte kan 'n gebrek aan kruisverwysings voorkom (UNESCO, 1975:3-7). Al hierdie faktore beïnvloed die prestasie van die stelsel.

Samevatting

Die invloed van die bogenoemde faktore op die IOHS kan nou samevattend beskryf word.

Die akkuraatheid waarvolgens terme gekies word, is die belangrikste kriterium wat die woordeskat van 'n indekseertaal raak, aangesien dit 'n invloed uitoefen op die presiesheid waarmee dokumente uit die IOHS herwin word.

Die vlak van uitputtendheid waarvolgens geïndekseer word, is die belangrikste kriterium wat die kwaliteit van 'n IOHS ten opsigte van herwinning bepaal. Hoe intensiewer 'n dokument geïndekseer word, hoe beter is 'n IOHS in staat om alle verbandhoudende dokumente wat die inligtingsnavraag verteenwoordig, te herwin. Hierdie herwinde dokumente is egter nie altyd almal bruikbaar nie, aangesien "te veel" onderwerpe baiekeer geïdentifiseer en geïndekseer is. Dit kan veroorsaak dat hierdie terme oor die potensiaal beskik om 'n vals koördinasie tydens herwinning te gee. Die teenoorgestelde is weer van toepassing waar uitputtendheid tot die minimum beperk word. Dáár kan 'n laer herwinning en hoër presiesheid bewerkstellig word deurdat die dokumente wat herwin word, absoluut verteenwoordigend van 'n navraag is. Deur net één term aan 'n onderwerp toe te ken, word die moontlikheid van vals koördinasies tot die minimum beperk (Lancaster, 1979:193-196).

Uit bogenoemde bespreking blyk dit dat die akkuraatheid waarvolgens terme gekies word en die vlak van uitputtendheid waarvolgens indeksering plaasvind, die prestasievermoë van 'n IOHS met betrekking tot herwinning en presiesheid bepaal. Die kwaliteit van indeksering kan gevolglik vasgestel word deur die IOHS te evalueer ten opsigte van herwinning en presiesheid.

Lancaster (1979:196-197) is van mening dat 'n onderskeid getref moet word tussen 'n afname in herwinning wat deur indekseerfoute veroorsaak word en 'n afname wat deur die vlak van uitputtendheid wat toegepas word, veroorsaak word. In die eerste geval word lae herwinning bewerkstellig as 'n indekseerder 'n onderwerp wat die kern van die inhoud van 'n dokument raak, nie identifiseer nie; in die tweede geval word 'n aspek van die onderwerp wat van belang is, nie in die indekseerproses verteenwoordig nie. In albei gevalle is dit moontlik dat sekere dokumente wat as hoogs relevant vir 'n soektog beskou kan word, nie herwin word nie.

Slotgedagte

Die volgende woorde van Soergel (1994:598) word as 'n slotgedagte aangehaal:

"Retrieval performance is a function of the agreement of the judgements of two parties, the indexer and the user, with respect to the viewpoints used in indexing, the general importance threshold used, and the specific judgements regarding the importance of a specific descriptor for a specific entity. The goal of indexing must be to maximize that agreement."

Vervolgens word die evaluering van 'n IOHS ondersoek om daaruit onder andere af te lei wat die kwaliteit van indeksering is.

4.3 Evaluering van die prestasievermoë van 'n IOHS

Geen indekseermetode kan as afgehandel beskou word as die kwaliteit van die indeks nie geëvalueer is nie. Een van die voordele van evaluering is dat dit 'n metode is wat gebruik kan word om gebruikerstevredenheid te waarborg. As gebruikers tevrede is, kan aanvaar word dat die kwaliteit en doeltreffendheid van die indekseerstelsel wat gebruik word, goed is. Lancaster (1968:118) sê ten opsigte van evaluering:

" . . . we analyse to determine how far the system is satisfying user requirements, to determine sources of system failure, and to determine how these failures may be best be remedied. "

Cleveland & Cleveland (1990:143) beweer dat dit die doel van evaluering is om te bepaal wat die effektiwiteit, geskiktheid en waarde van indekseerterme is wat aan dokumente toegeken is. Volgens Soergel (1985:126) moet egter tussen toetsing en evaluering onderskei word. Hy is van mening dat toetsing op die vasstelling van prestasievermoëns dui, soos byvoorbeeld herwinning, terwyl evaluering 'n waarde aan die prestasievermoë van die IOHS toeken soos dit in 'n bepaalde situasie geld. Die kwaliteit van indekseerterme bepaal in 'n groot mate die kwaliteit van die IOHS as 'n geheel.

In literatuur oor evaluering is die meeste outeurs dit eens dat daar nog geen metode vir evaluering is waaroor algemeen saamgestem word nie. Die meeste meeteenhede wat in evaluering gebruik word, is kwantitatief van aard. Dit is egter noodsaaklik dat evaluering kwaliteit én kwantiteit moet bepaal. Uit kwantitatiewe resultate kan egter afleidings gemaak word wat 'n aanduiding van kwaliteit kan gee. Twee meeteenhede wat veral gebruik kan word om kwalitatiewe afleidings te maak, is herwinning en presiesheid. (Vergelyk 4.4.)

Ander meeteenhede wat in 'n mindere mate 'n aanduiding van die kwaliteit van die stelsel gee, is byvoorbeeld:

- dekking, dit wil sê hoeveel dokumente in die versameling teenwoordig is wat 'n sekere onderwerp dek, asook die aard van die dokumente;
- aktualiteit, dit wil sê die persentasie nuwe dokumente wat 'n gebruiker uit 'n IOHS herwin wat die eerste keer onder sy aandag kom, dit wil sê nuwe inligting;
- resenteheid, dit wil sê of die nutste inligting in die versameling beskikbaar is;
- akkuraatheid van indeksering, dit wil sê hoe akkuraat die indeksering wat toegepas is, die inhoud van 'n dokument weerspieël;

- reaksietyd, dit wil sê die hoeveelheid tyd wat dit 'n gebruiker neem vanaf die formulering van 'n navraag tot die ontvangs van die soekresultate;
- duplisering, die vermoë wat die stelsel het om duplikaatdokumente of dokumente waarvan die gebruiker reeds bewus is en wat hy/sy reeds gesien het, uit te skakel.

Vir hierdie studie word herwinning en presiesheid as meeteenhede uitgesonder, want dit oefen die grootste invloed op die prestasievermoë van die stelsel en gevolglik ook op gebruikerstevredenheid uit.

4.4 Herwinning en presiesheid as meeteenhede

Met herwinning word die persentasie relevante dokumente bedoel wat uit die totale beskikbare relevante dokumente in die versameling herwin word. Herwinning meet die wyse waarvolgens die IOHS in staat is om relevante dokumente te herwin. Presiesheid daarteenoor is die aantal relevante dokumente uit die totale hoeveelheid wat tydens 'n soektog herwin word; dus, presiesheid bepaal die vermoë van die IOHS om nierelevante dokumente terug te hou.

Herwinning en presiesheid kan soos volg bereken word:

Herwinning:

$$\frac{\text{aantal relevante dokumente wat herwin is}}{\text{die totale aantal relevante dokumente in die versameling}}$$

Gestel 30 dokumente word oor hartsiektes herwin en dat daar 50 dokumente in die versameling oor hartsiektes is. Dan word herwinning soos volg bereken: $30 \div 50 = 60\%$ van die totale aantal relevante dokumente in die versameling word herwin.

Presiesheid:

$$\frac{\text{aantal relevante dokumente wat herwin is}}{\text{totale aantal dokumente wat herwin is uit die versameling}}$$

Gestel met 'n soektog oor depressie word 80 dokumente herwin, maar slegs 60 is relevant. Dan word presiesheid soos volg bereken: $60 \div 80 = 75\%$ van die dokumente wat herwin is, is relevant ten opsigte van die onderwerp.

Die resultate wat met hierdie twee meeteenhede verkry word, is bepalend vir 'n IOHS se bruikbaarheid, aangesien dit 'n aanduiding gee van hoeveel relevante dokumente die stelsel deurlaat - dit is wat gebruikers van 'n stelsel verwag (Cleveland, 1990:149).

Relevansie

In die bepaling van herwinning en presiesheid hierbo is gesien dat relevansie telkens as kriterium gebruik word. Dit is baie moeilik om te bepaal wat relevant is en wat nie. "Relevansie as sodanig kan nie gemeet word nie, en is ook nie 'n meeteenheid nie", bevestig Boon (1976:84). Soergel (1985:127) se mening is: "Any computations of recall, discrimination, and so on are only as reliable as the relevance judgements on which they are based."

Relevansie wissel van gebruiker tot gebruiker. 'n Dokument wat vir een persoon relevant is, is nie noodwendig vir 'n ander een relevant nie. Dit hang af van die persoon se ervaring en kennis van 'n onderwerp. By evaluering word die term relevansie gebruik om die verwantskap wat daar tussen 'n dokument en die navraag bestaan, uit te druk. "Relevansie kan beskou word as die toestand van 'n bepaalde dokument ten opsigte van 'n gebruiker se navraag, op grond waarvan meting gedoen kan word" (Boon, 1976:84). Hierdie verwantskap kan egter nie presies afgebaken word nie, aangesien dit vir elke individu sal verskil. In die literatuur kom daar verskillende menings oor relevansie na vore.

Park (1994:135-136) sê dat 'n gebruiker die belangrikste aspek is wat in ag geneem moet word by die vasstelling van relevansie. Relevansie moet vanuit die gebruiker se oogpunt onlosmaaklik verbind wees aan die inligtingsbehoefte wat deur hom geïdentifiseer is. Sy wys daarop dat relevansie nie met die onderwerp van die navraag of die situasie waarin die gebruiker hom bevind, verband hou nie. Om relevansie sinvol af te baken, moet aandag aan die volgende sake gegee word: die doel waarvoor inligting

benodig word; hoe dit in ooreenstemming gebring kan word met inligting wat alreeds in die gebruiker se geheue aanwesig is; die stadium van die navorsing van die individu; ten slotte, die hoeveelheid tyd wat beskikbaar is om 'n soektog uit te voer.

Barry (1994:149-150) voeg hierby dat die indekseerterme van die IOHS wat gebruik word met die navraag wat deur die gebruiker geformuleer word, verenigbaar moet wees.

Rees & Saracevic (1966:225-234) beklemtoon die belangrikheid van gebruikersgeoriënteerdheid, die subjektiwiteit van relevansie en die ooreenstemming daarvan met die inligtingsbehoefes van 'n individu wat altyd in 'n veranderlike staat verkeer. Hulle wys ook daarop dat interpretasies van 'n dokument nooit identies sal wees nie, aangesien elke individu se kognitiewe verwerking van die inhoud sal verskil. Daarom moet relevansie eerder as 'n vergelykende begrip beskou word en in ooreenstemming met die kognitiewe proses gebring word.

Samevattend kan gesê word dat relevansie beskou kan word as die logiese antwoord (respons) op 'n gestelde navraag, dit wil sê soos die gebruiker dit verwoord het. Relevansie behoort objektief te wees wanneer dit as kriterium by die prestasie van 'n IOHS gebruik word.

Pertinensie

Vir die doel van hierdie bespreking is dit ook nodig om tussen *pertinensie* en *relevansie* te onderskei. Pertinensie het te doen met die toepaslikheid van inligting op 'n bepaalde tydstip, in 'n bepaalde situasie. Die gebruiker alleen kan besluit watter inligting vir hom pertinent is. Die waarde wat die gebruiker aan die inhoud van 'n dokument heg, word in ooreenstemming met 'n inligtingsbehoefte wat op dieselfde tydstip ondervind word, gebring. Die waarde wat aan die dokument toegeken word, is net van toepassing op 'n sekere situasie en tydens 'n sekere tydstip waarin 'n gebruiker hom bevind. Dieselfde dokument kan bes moontlik as nutteloos beskou word op 'n ander tydstip wanneer die gebruiker se inligtingsbehoefes nuwe vereistes stel (Lancaster, 1979:264).

Soergel (1985:128) maak 'n onderskeid deur na onderwerpsrelevansie ("topical relevance") te verwys as dokumente wat ongeag die gebruiker se agtergrond en omstandighede as relevant beoordeel word, teenoor toepaslikheid ("appropriateness") wat na dokumente verwys wat vir die gebruiker se agtergrond en huidige situasie geskik is. As 'n dokument terselfdertyd onderwerpsrelevant én toepaslik vir 'n gebruiker se agtergrond is, word dit pertinente dokumente genoem.

Relevansie en pertinensie kan egter nie in kompartemente ingedeel word nie, aangesien die begrippe veranderlike elemente bevat, soos blyk uit bogenoemde bespreking - "This is an arbitrary decision made purely for the sake of convenience" (Lancaster, 1979:264).

Gebruikersverwagtinge

Soos reeds gemeld, moet die IOHS in 'n gebruiker se inligtingsbehoefte voorsien deur vir hom voldoende relevante dokumente te herwin. Daar is volgens Fourie (1990:141) 'n paar faktore wat die gebruikers se verwagtinge met 'n inligtingsoektog kan beïnvloed:

- te veel bibliografiese verwysings word herwin wat te lank neem om deur te werk;
- die bibliotekaris kan nie werklik die inligtingsbehoefte van die gebruiker bepaal nie;
- die inligtingstelsel herwin verwysings in plaas van dokumente of inligting wat addisionele tyd verg om die oorspronklike dokumente op te spoor;
- die inligting wat aan die gebruiker verskaf word, is relevant maar nie altyd bruikbaar nie, omdat dit soms in 'n vreemde taal is of slegs op microfiche beskikbaar is.

As bogenoemde omstandighede voorkom, kan die gebruiker die soektog as 'n mislukking beskou en dink dat die prestasie van die IOHS laag en swak is.

Samevatting

Herwinning en presiesheid as meeteenhede kan gebruik word om kwalitatiewe afleidings te maak. Al word die evaluering kwantitatief gedoen (byvoorbeeld die hoeveelheid relevante dokumente), is dit 'n aanduiding van die kwaliteit van die IOHS. As herwinning en presiesheid hoog is, dui dit op die vermoë van die stelsel om relevante dokumente te herwin en nierelevante dokumente terug te hou - faktore wat vir die gebruikers belangrik is.

Die kwaliteit van die stelsel word in 'n groot mate deur die kwaliteit van die indekseerterme bepaal. As daar met behulp van herwinning en presiesheid aangedui kan word dat die kwaliteit van die stelsel goed is, kan daaruit afgelei word dat die kwaliteit van die indekseerterme ook goed is. Dit is 'n stap in die verwesenliking van die doel van die studie, naamlik om vas te stel wat die kwaliteit van toegekende terme teenoor afgeleide terme is.

Vervolgens sal daar gekyk word watter invloed die indekseerterme op die prestasie van die IOHS het. Soos reeds in 2.4.1 gesien, behoort die indekseertaal die kommunikasieproses tussen 'n gebruiker en die bron van inligting te bevorder. As hierdie kommunikasie nie glad verloop nie, kan die inligtingsbehoefte van die gebruiker nie bevredig word nie. Die kwaliteit van die indekseerterme is dus hier van groot belang.

4.5 Die invloed van die indekseerterme op herwinning en presiesheid

4.5.1 Die invloed van spesifiekheid van terme

Met die spesifiekheid van terme word volgens Aluri et al. (1991:64) die wyse bedoel waarvolgens die regte "woord" aan 'n geïdentifiseerde konsep toegeken word. Lancaster (1986:138-139) is van mening dat 'n gebrek aan spesifiekheid in die woordeskaf van 'n IOHS veroorsaak dat die verwantskappe tussen terme verwarrend en vals aangedui word. As die woordeskaf nie in staat is om die presiese konsep uit te druk nie, sal dit tot gevolg hê dat die navraag van 'n gebruiker ook 'n gebrek aan presiesheid toon en sodoende die relevansie daarvan benadeel. Soergel (1985:330) verwys na spesifiekheid as die generiese vlak waarvolgens konsepte toegeken word om

'n entiteit te beskryf. Cleveland & Cleveland (1983:181) sê dat die akkuraatheid van indeksering deur die mate van spesifiekheid beskryf word, naamlik in watter mate die deskriptor wat gekies word om 'n spesifieke konsep te verteenwoordig, met die presiese betekenis van laasgenoemde ooreenstem. Die spesifiekheid van terme word deur die volgende faktore bepaal: die tipe en grootte versameling wat geïndekseer word, die gebruikersprofiel wat van die versameling gebruik maak en die hoeveelheid terme wat beskikbaar is om 'n onderwerp te beskryf. Die term *bome* is byvoorbeeld heeltemal spesifiek genoeg vir 'n klein biblioteek wat slegs 'n paar dokumente oor bome besit, terwyl ook meer spesifieke terme soos *dennebome*, *akkerbome*, *eikebome* vir 'n meer gespesialiseerde biblioteek wat baie publikasies oor bome besit, nodig sal wees. 'n Instansie soos Medline funksioneer op dieselfde beginsel deur 'n onderwerp altyd onder die mees spesifieke term te indekseer. Indien 'n onderwerp oor sonbrand handel, word alle verwante onderwerpe onder "sunburn" geïndekseer en nie onder "burns" nie. Terme wat gekies word, moet so spesifiek moontlik wees vir die behoeftes van die instansie en die inligtingstelsel wat gebruik word. Die *UNISIST Indexing Principles* beklemtoon ook dat die indekseerder die waarde wat hy aan 'n spesifieke konsep heg, in ooreenstemming met die behoeftes van gebruikers en die grootte van die versameling moet bring, anders kan terme te spesifiek of te algemeen gekies word (UNESCO, 1975:7). Lancaster (1986:140) wys daarop dat die spesifiekheid waarvolgens terme gekies word die bepalendste faktor by herwinning is, want geen verwysings sal herwin kan word as terme nie spesifiek genoeg gekies is nie. Die spesifiekheid van terme het dus 'n direkte invloed op die geslaagdheid van die navraagproses.

Die presiesheid waarvolgens dokumente herwin word, dit wil sê hoe relevant die dokument is, word dus bepaal deur die spesifiekheid waarvolgens terme toegeken word. Lancaster (1986:140) wys op die belangrike beginsel dat hoe meer betekenis daar aan 'n konsep geheg kan word, hoe meer onderwerpskennis van die indekseerder vereis word, aangesien die totale aantal terme al hoe meer gespesialiseerd en meer word. By rekenaaraafgeleide indeksering waar terme nie beheer word nie, stem die terme dalk meer met natuurlike taal ooreen. Die nadeel is egter dat 'n soektog nie binne 'n beheerde raamwerk gekonstrueer kan word soos wat die geval met

toegekende indeksering is nie. 'n Gekontroleerde woordeskat beheer sinonieme en naby sinonieme en skakel verwante terme aan mekaar deur middel van 'n raamwerk van hiërargiese en ander skakelverwysings wat riglyne aan 'n gebruiker tydens die soek van inligting verskaf (Lancaster, 1986:210). Aitchison & Gilchrist (1987:23) deel hierdie siening deur te sê dat die spesifiekheid van die herwinningswoordeskat grootliks afhanklik is van die indekseerterme wat die onderwerp in diepte en detail moet uitdruk. Hulle wys ook daarop dat die omvattendheid waarvolgens 'n onderwerp geanaliseer word, 'n invloed op die akkuraatheid kan uitoefen waarvolgens konsepte geïdentifiseer word en op die geslaagdheid om nierelevante dokumente uit 'n soektog uit te sluit.

Lancaster wys ook daarop dat die tweede grootste faktor wat stelselprestasie beïnvloed, die verwarrende of vals betekenis is wat aan 'n term geheg word. Dit kan tot gevolg hê dat vals koördinasies en verkeerde termverwantskappe tydens 'n soektog na vore kom. Dit gebeur as 'n woordeskat nie daartoe in staat is om alle verwante terme as 'n eenheid deur middel van 'n hiërargiese struktuur en/of 'n kruisverwysingsnetwerk saam te voeg nie. Spesifiekheid word kernagtig soos volg deur Lancaster (1986:145-146) saamgevat:

"Vocabulary specificity directly controls the size of the classes into which documents are placed. The size of the classes defined by a vocabulary is much more important than the arrangement of these classes."

4.5.2 Die invloed van 'n beheerde woordeskat

Die essensie van 'n beheerde woordeskat is dat één term vir 'n aantal verwante of naby verwante konsepte gekies word en dat daar van ander terme na die term wat gekies is, verwys word. Cleveland & Cleveland (1990:80-81) wys daarop dat 'n beheerde woordeskat se invloed by dié punt begin - hier waar alle verwante terme byeengebring word sodat terme nie deur die indeks onder verskillende sinonieme of variante vorme versprei word nie. Dit bring ook twee konsepte ter sprake, naamlik "literary warrant" en "user warrant" wat onderskeidelik met literêre volmag - of selfs tekstuele volmag - en gebruikersvolmag vertaal kan word. Eersgenoemde hou in dat 'n konsep wat baie in die literatuur voorkom, gebruik kan word om 'n gebruiker na relevante

dokumente vir sy inligtingsbehoefes te lei. Dienooreenkomstig impliseer gebruikersvolmag dat 'n term wat in die mond van die gebruiker leef, as 'n leidraad in die keuse van 'n beheerde term sou kon dien. As daar reeds 'n term gekies is, kan dit die behoefte deur 'n kruisverwysing onder die aandag bring. Alhoewel die terme in 'n beheerde woordeskat nie soveel vryheid aan die gebruiker verskaf soos wat die geval met natuurlike taal is nie, bring die konsekwente gebruik daarvan mee dat herwinning akkuraat kan plaasvind.

'n Beheerde woordeskat verskaf dus aan die gebruiker 'n gestandaardiseerde woordeskat wat gebruik word om sinonieme en naby sinonieme te beheer en op die manier konsekwentheid te bewerkstellig. Verwarring onder die gebruikers word tot 'n minimum beperk deurdat slegs een voorkeurterm gebruik word om 'n sekere konsep uit te druk. Die terme is ook so gekies dat 'n onderskeid getref word tussen terme wat verwarrend kan wees, soos in die geval van homonieme. Kruisverwysings en ander skakelverwysings word as voorsorgmaatreël gebruik om die horisontale en vertikale verwantskappe tussen terme aan die gebruiker uit te wys. Die terme wat deur 'n beheerde woordeskat beskikbaar gestel word, word deur middel van voorafbesluitneming en kontrole tot betekenisvolle eenhede gebind, sodat vals koördinasies tot die minimum beperk word. 'n Beheerde woordeskat verskaf aan 'n gebruiker 'n soekstrategie oor hoe om selektiewer of omvattender tydens die herwinning van inligting te werk te gaan. Die gebruiker word voorsien van "onbewustelike" hulp deurdat addisionele terme gesuggereer word (Lancaster, 1986:147,151).

'n Beheerde woordeskat se waarde lê egter daarin dat die probleme en verwarring rondom natuurlike taal beheer kan word. In sekere vakgebiede, naamlik in die geesteswetenskappe, is vakterminologie nie so vas soos byvoorbeeld in die natuurwetenskappe nie, en kan daar verwarring en dubbelsinnigheid in die terme voorkom. 'n Beheerde woordeskat se belangrikste funksie is om 'n lys van woorde beskikbaar te stel wat inligting dadelik binne die bereik van die gebruiker plaas. Hierdie lys woorde het 'n direkte invloed op herwinning deurdat sinonieme beheer word en afhange van die stelsel wat gebruik word, 'n aanduiding na breër of meer spesifieke terme gegee word.

Lancaster (1986:161) vat die voordele van 'n beheerde woordeskat soos volg saam:

"It tends to reduce semantic ambiguities, to promote consistency in the representation of the subject matter, and to facilitate the conduct of comprehensive searches."

Samevatting

Die kwaliteit van indeksering en indekseerterme is belangrik, want dit het 'n direkte invloed op die prestasie van 'n IOHS en gevolglik ook op gebruikerstevredenheid. Die gebruiker moet in staat gestel word om met min moeite relevante inligting deur die gebruik van soekterme te herwin. Aan die ander kant moet die IOHS daarin slaag om nierelevante inligting terug te hou.

HOOFSTUK 5

'N EVALUERING VAN TOEGEKENDE INDEKSERING TEENOR AFGELEIDE INDEKSERING

5.1 Inleiding en probleemstelling

Daar is aangetoon dat die prosesse van toegekende indeksering en afgeleide indeksering wesentlik van mekaar verskil. In die een speel menslike logika en redenasie 'n belangrike rol, in die ander glad nie. In hierdie hoofstuk gaan gekyk word OF daar merkbare verskille in die herwinningsprestasie van die twee verskillende metodes is en ook of dit die presiesheid van soekresultate affekteer. Vir hierdie doel is inligtingsoektogte in verskillende databasisse uitgevoer, en in elke soektog is toegekende terme sowel as afgeleide terme gebruik. Deur die resultate van elke soektog met mekaar te vergelyk, kan afleidings gemaak word oor watter tipe indekseerterme die beste presteer ten opsigte van die hoeveelheid dokumente wat herwin word. Verder is die relevansie van die herwinde dokumente ontleed om af te lei hoe die terme met betrekking tot presiesheid by herwinning presteer.

Vooraf is dit nodig om daarop te wys dat afleidings en gevolgtrekkings wat uit sulke resultate gemaak word, nie absoluut is nie en net 'n algemene aanduiding van die prestasie gee. In die literatuur word verskillende standpunte oor die prestasie van elk van die tersaaklike indekseermetodes gehuldig. Blair en Maron (1985:289-299) het byvoorbeeld bewys dat die presiesheidsvlak van afgeleide indeksering ongeveer 79% korrek was tydens veertig soektogte wat in 'n regsdatabasis uitgevoer is, terwyl die herwinningspersentasie slegs ongeveer 20% was. Daarteenoor is Hartley (1990:168) van mening dat afgeleide indeksering in sommige gevalle nie 'n hoë vlak van presiesheid het nie. Redes daarvoor is dat sommige onderwerpsareas terme gebruik wat nie baie spesifiek is nie en dat 'n inligtingsvaardigheid van die gebruiker vereis word om korrekte taaltoepassings te maak. Studies wat in die verlede gedoen is, het bewys dat daar met die een metode dokumente herwin wat nie met die ander een moontlik is nie. In hierdie verband sê Svenonius (1986:331-340) dat 'n goed

gekombineerde tegniek van albei metodes ontwikkel behoort te word om 'n hoë vlak van herwinning en presiesheid te bewerkstellig.

Die term evaluering impliseer in hierdie geval dat die kwaliteit (of waarde) wat toegekende en afgeleide terme het, beoordeel sal word. Die kwaliteit van die terme kan egter nie volgens 'n absolute skaal gemeet word nie en die toepaslikheid daarvan in die situasie moet in aanmerking geneem word. 'n Belangrike feit om in die uitvoering van 'n evalueringseksperiment soos hierdie in gedagte te hou, is die eksterne waarde waaroor dit beskik, naamlik in watter mate die aspekte wat geëvalueer word en die resultate wat verkry word, verteenwoordigend van die teikensituasie is waarop gekonsentreer word. 'n Verdere oorweging is of die resultate wat verkry word, op die werklike situasie toegepas kan word. Eksperimentele bevindinge soos hierdie moet geëvalueer word ooreenkomstig die toepaslikheid van die data, die wetenskaplike belang daarvan en in watter mate die data veralgemeen kan word (Busha & Harter, 1980:43). Vir die omvang van hierdie studie word die afleidings wat gemaak word, tot 'n algemene stelling beperk, naamlik dat die vergelykende herwinningspresentasie by elke soektog 'n aanduiding kan wees van die kwaliteit van die indekseerterme en dat laasgenoemde ook die presiesheid van die IOHS beïnvloed.

Frost (1989:165-179) het 'n studie soortgelyk aan hierdie uitgevoer om die volgende te bepaal: eerstens, die ooreenkoms tussen toegekende Library of Congress-saakhoofde en afgeleide terme uit titels in bibliografiese rekords; tweedens, die implikasies daarvan vir onderwerpstoegang in 'n intydse katalogus. Sy het bevind dat gemiddeld 20% van die woorde wat in die titel en die ekserp teenwoordig is, dieselfde is as woorde wat in die tesourus voorkom. Sy wys ook daarop dat die meeste van hierdie studies wat uitgevoer is nie daarop toegespits is om die verskil in die herwinning tussen toegekende en afgeleide terme te evalueer nie, maar om te bepaal in watter mate die menslike indekseerbesluite wat uitgevoer word, suksesvol deur 'n rekenaar nageboots kan word.

5.2 Navorsingsontwerp

In 4.1 is na Lancaster se mening verwys dat indeksering net ex post facto geëvalueer kan word. In Latyn beteken dit letterlik na die feit of retrospektiewelik. 'n Ex post facto-studie is nie 'n eksperimentele ontwerp nie, aangesien kontrole nie uitgeoefen kan word nie. So 'n studie is nie toegespits op die bekendstelling van onafhanklike veranderlikes en die kontrolering van ander veranderlikes nie. In 'n ex post facto-studie word die klem geplaas op die analisering van dit wat alreeds gebeur het. Dit is 'n poging om die aanleiding tot die situasie van die situasie te isoleer (Busha & Harter, 1980:43).

Die soektogte in hierdie studie is dus in bestaande databasisse en indekse uitgevoer, eerstens met toegekende terme in die onderwerpsveld en tweedens met woorde in die titel (titelafgeleide terme). Die herwinning wat elkeen van die twee metodes lewer, word telkens uit die soekresultate vasgestel, terwyl die titel en die ekserp ontleed word om uit die relevansie die presiesheid vas te stel.

5.3 Seleksie van databasisse

Die databasisse wat gebruik is, is gekies uit die lys van beskikbare databasisse in die Ferdinand Postma Biblioteek. Dit is baie wyd gekies sodat dit verteenwoordigend van verskillende vakrigtings sowel as programpakkette kan wees. Al hierdie databasisse het egter ooreenstemmende kenmerke, byvoorbeeld dat dit gebruik maak van deskriptore of trefwoorde én rekenaarmatig beskikbaar is. Verder beskik hulle oor 'n soekfasiliteit wat die gebruiker in staat stel om in die onderwerpsveld sowel as in die titelveld te soek.

Vervolgens word die databasisse wat gebruik is, bespreek.

5.3.1. Ferdinand Postma Biblioteek se intydse katalogus

Die Ferdinand Postma Biblioteek se intydse katalogus sluit die totale versameling boeke, tydskrifte en ander media in die Biblioteek in. Ontsluiting van materiaal word gedoen volgens die gestandaardiseerde reëls van AACR2, terwyl die toekenning van indekseerterme volgens die Library of Congress Subject Headings gedoen word. Hierdie terme word in Afrikaans vertaal deur van standaardwoordeboeke en -onderwerpslyste gebruik te maak. Die Engelse terme word met ontsluiting behou en verskaf dienooreenkomstig addisionele soekpunte.

5.3.2 In U Lig-databasis

Die In U Lig-databasis word intern deur personeel van die Ferdinand Postma Biblioteek opgestel - met die doel om Christelike wetenskaplike en kerklike tydskrifte wat nie in ander databasisse geïndekseer word nie, te ontsluit. In hierdie databasis word hoofsaaklik van natuurliketaal-indeksering gebruik gemaak om die inhoud van dokumente te ontsluit. Die indekseerterme word egter gekontroleer deur dit in 'n trefwoordlys beskikbaar te stel. Indekseerreëls word nie presies in die keuse van terme toegepas nie. *Sondag* en *Sabbat*, *Duiwel* en *Satan* word byvoorbeeld alternatiewelik gebruik.

5.3.3 Repertorium van Suid-Afrikaanse tydskrifartikels

Die Staatsbiblioteek hou die Repertorium van Suid-Afrikaanse tydskrifartikels in samewerking met die Johannesburgse Openbare Biblioteek in stand. Die Repertorium word opgebou uit indekseerinskrywings op meer as 460 Suid-Afrikaanse tydskrifte wat volledig geïndekseer word. Algemener tydskrifte wat nie so gespesialiseerd soos vakwetenskaplike tydskrifte is nie, word geïndekseer. Die onderwerpshoofde wat toegeken word, is gebaseer op Library of Congress Subject Headings en Medical Subject Headings wat internasionaal aanvaarbare lyste is en internasionale standaardisering van terminologie teweeg bring. Die Repertorium maak dus gebruik van 'n

beheerde woordeskat en wanneer die woordeskat nie in die Suid-Afrikaanse behoeftes voorsien nie, word van eilandse terme gebruik gemaak.

5.3.4 ERIC

Die Educational Resources Information Center (ERIC) is 'n internasionale inligtingstelsel wat ontwerp is met die doel om direkte toegang tot opvoedingsverwante literatuur te verskaf. Hierdie databasis word geldelik deur die Amerikaanse regering se Departement van Opvoeding in stand gehou en bevat meer as 700 000 bibliografiese rekords van dokumente en tydskrifartikels. ERIC word op datum gehou deur 16 gedesentraliseerde sentrums wat oor opvoedkundige kennis beskik en wat hoofsaaklik by universiteite of ander professionele verenigings gesentreer is. Instandhouding word ook verder deur ander ERIC Adjunct Clearinghouses gedoen. Hierdie sentra identifiseer bruikbare opvoedkundige inligting in onderwerpsvelde soos lees en kommunikasievaardighede, wetenskap en wiskunde, asook sosiale studies en ander velde, soos sekondêre of elementêre onderwys in landelike en stedelike gebiede. Elke item word gekatalogiseer, geabstraheer en geïndekseer volgens die sleutelwoorde van 'n beheerde woordeskat - die Thesaurus of ERIC Descriptors. Inligting kan deur die ERIC-databasis herwin word. ERIC het twee fundamentele indekseerreëls, naamlik:

- net dit wat in die dokument voorkom, word geïndekseer;
- indeksering vind volgens die vlak van spesifiekheid in die dokument plaas.

5.4 Metode van ondersoek

Die metode van ondersoek wat gevolg is om die prestasie van toegekende en afgeleide indeksering vas te stel, word kortliks beskryf. 'n Aantal beskrywende terme wat 'n onderwerp verteenwoordig, is gekies. Elkeen van die terme is telkens as 'n soekpunt gebruik om tydskrifartikels oor die onderwerp op te spoor. Die terme is eers in die onderwerpsveld gesoek (wat die toegekende terme verteenwoordig) en daarna in die titelveld (wat die

rekenaar afgeleide terme verteenwoordig): in die intydse katalogus van die Ferdinand Postma Biblioteek is terme in die onderwerpsveld én titelveld gesoek; in die In U Lig- en Repertorium-databasisse is daar in die trefwoordveld sowel as die titelveld gesoek; ten slotte, in die ERIC-databasis volgens die Titlewords en ERIC Subject Headings. As soekterme is sover moontlik enkelvoudige woorde gebruik, sodat dieselfde woorde in die titelveld en die onderwerpsveld gesoek kon word. Daar is geredeneer dat die kans dat 'n enkelwoord in 'n titel sal voorkom, groter is as wat die geval sou wees met saamgestelde terme.

Die ondersoek is informeel in die Naslaandepartement van die Ferdinand Postma Biblioteek uitgevoer. Die onderwerpe wat gekies is, kom uit verskillende vakrigtings, sodat die resultate sal varieer. Die verskillende soekresultate word volgens 5.5 as beskrywende statistiek in tabelvorm weergegee en in 5.6 ontleed. Tydens die soekproses is die volgende stappe uitgevoer:

- die kies van 'n databasis;
- die bepaling en vastelling van die onderwerp waarvoor bibliografiese verwysings herwin moet word;
- die seleksie van enkelvoudige terme;
- die uitvoering van die soektog;
- 'n skermvertoon van soekresultate en die uitdruk daarvan.

5.5 Soekresultate

5.5.1 Resultate wat die herwinningsprestasië van titelafgeleide en toegekende indeksering aandui

Die terme (woorde) waaronder gesoek is, word in kolom een gegee. In kolom twee word die aantal verwysings wat in die titelveld gevind is, gegee. Hier sal die woord(e) in die titel as soekpunt(e) gebruik wees (titelafgelei). In kolom drie word die aantal verwysings wat in die onderwerpsveld gevind is, gegee. Hier sal die soekterme deskriptore of saakhoofde wees (toegeken).

SOEKTERM	TITELVELD	ONDERWERPSVELD
FERDINAND POSTMA BIBLIOTEEK INTYDSE KATALOUS		
<i>air pollution</i>	4	6
<i>academic achievement</i>	0	5
<i>skin absorption</i>	0	3
<i>nasionalisme</i>	17	47
<i>draaiboeke</i>	0	12
<i>mongolisme</i>	0	12
REPERTORIUM		
<i>boats</i>	4	12
<i>astronomy</i>	8	77
<i>disabled</i>	46	196
<i>Indians</i>	15	15
<i>child abuse</i>	12	110
<i>abortion</i>	24	67
IN U L I G		
<i>genadedood</i>	5	19
<i>doodstraf</i>	11	12
<i>aborsie</i>	12	31
<i>skepping</i>	21	125
<i>doop</i>	20	29
<i>dorslandtrek</i>	1	1
ERIC		
<i>planetariums</i>	0	20
<i>regional dialects</i>	0	157
<i>disease control</i>	4	578
<i>eating disorders</i>	83	111
<i>vikings</i>	0	3
<i>unidentified flying objects</i>	1	3

5.5.2 Resultate wat die presiesheid van herwinde dokumente aandui

Hier word eerstens die term gegee waaronder gesoek is, dan titels van dokumente wat met daardie soekterm herwin is en laastens die saakhoofde/deskriptore wat aan die dokument toegeken is. Dit is voorbeelde van toegekende indeksering. Daarna word titels gegee wat die resultate van 'n soektog onder afgeleide terme is.

FERDINAND POSTMA BIBLIOTEK INTYDSE KATALOGUS

Soekterm: *Academic achievement*

Voorbeelde van resultate uit onderwerpsveld:

Outeur:	Purkey, William Watson
Titel:	Self concept and school achievement
Onderwerpe:	Academic achievement Self-perception Akademiese prestasie Selfwaarneming

Outeur:	McLeod, John
Titel:	Fostering academic excellence
Onderwerpe:	Gifted children--Education Academic achievement Creative thinking Begaafde kinders--Onderwys Kreatiewe denke (Onderwys)

Outeur:	Lindgren, Henry Clay
Titel:	The psychology of college success: a dynamic approach
Onderwerpe:	College student orientation College students--Psychology Academic achievement Study skills Prediction of scholastic success Universiteitsomgewing Leerproses, Psigologie van die Studente aanpassing

Voorbeelde van resultate uit die titelveld:

Academic acceleration of gifted children / edited by W. Thom

Academic achievers and non-achievers : an analysis of some F

Academic ambitions and library development : the American Bu

Academic and legal deposit libraries : an examination guide

Soekterm: *Mongolisme*

Voorbeelde van resultate uit onderwerpsveld:

Outeur:	Steyn, I.N. (Izak Nicolaas)
Titel:	Die opleibaarheid van die kind met Downsindroom
Onderwerpe:	Mongolisme Verstandelik gestremdes-Onderwys

Titel:	Gids vir genetiese afwykings / Genetiese Dienste, Departement van Gesondheid, Welsyn en Pensioe
Onderwerpe:	Genetiese afwyking Fetale alkoholsindroom Bloedgroepe Dooftheid Mongolisme Hemofilie Hiperlipidemia Spina bifida Tay-Sachs-siekte Teratogenese Sisti-fibrose

Outeur:	Grundlingh, Maria Elizabeth
Titel:	Downsindroom of mongolisme
Onderwerpe:	Mongolisme Rorschachtoets

Voorbeeld van resultate uit titelveld:

Mongolentzeit / Von Bertold Spuler

Mongolistik / Mit beitragen von Nikolause Poppe . . . [et al.]

Mongols

Mongols / David Morgan

Monica.

Monist monograph ; no. 1

ERIC

Voorbeeld van resultate uit onderwerpsveld:

Soekterm: *Vikings*

*Titel:	The Viking Discovery : 1,000 years ago
Outeur:	Alison, S R
Tydskrif:	Instructor, v06 n3 p46-48 Oct 1986
Ekserp:	Suggestions for a study unit for intermediate grades on the Viking discovery of North America are offered. Class activities and resources are listed.
Deskriptore:	Class Activities History Instruction Intermediate Grades Land Settlement North American History
Identifiseerders:	Vikings

Titel:	Reviews
Outeur:	Greenleaf, Floyd; and Others
Tydskrif:	History Teacher, v19 n2 p303-14 Feb 1988
Ekserp:	Reviews eight textbooks, readers and books. Topics include Latin America, colonial America, the Carolinians, women in French textbooks, the Vikings, the Soviet Union, nineteenth-century, Black America, and Ernest Rutherford.
Deskriptore:	<ul style="list-style-type: none"> Ancient History Black History Black Studies Book reviews Civil War (United States) Colonial History (United States) Feminism Latin American History Literary Criticism Textbook Evaluation Women's Studies World history
Identifiseerders:	<ul style="list-style-type: none"> Russia Rutherford (Ernest) Vikings

Die bibliografiese rekord wat met 'n * aangedui is, is uit die titelveld herwin.

Soekterm: *Planetarium*

Titel:	The perfect field trip for International Space Year
Outeur:	Hofman, Helenmarie
Tydskrif:	Science Scope, v15 n5 p30-31 Feb 1992
Ekserp:	Provides a poster containing activities that are part of the International Space Year. Describes activities in which students visit a local planetarium and re-create a National Aeronautic and Space Administration (NASA) space flight using reports, articles, NASA slides, and videos. Lists organizations, publications, and other resources related to planetariums use.
Deskriptore:	Class activities Enrichment activities Field trips Intermediate Grades Junior High Schools Middle Schools Planetariums Resource materials Science activities Science education Science instruction Space Exploration Space Sciences
Identifiseerders:	National Aeronautics and Space Administration Posters

*Titel:	A planetarium in the physics room
Outeur:	Payne, Mark M
Tydskrif:	Physics Teacher, v22 n8 p518-19 Nov 1984
Ekserp:	Describes how a planetarium was constructed on the ceiling of a traditional physics laboratory. Similarities and differences between the "poor man's" planetarium and other planetariums are noted.
Deskriptore:	Astronomy High Schools Physics Planetariums Science Education Science laboratories Secondary School Science

REPERTORIUM

Soekterm: *Boats*

Voorbeelde uit die onderwerpsveld:

Titel:	Sandock-Austral Group
Tydskrif:	Armed forces(1987);p.13-16 Apr.
Ekserp:	Discusses the Sandock-Austral Group and its three divisions, as well as one of its vehicles and boats
Trefwoorde:	Military engineering Technology Transportation Military vehicles Missile boats Etand Ratel Uklebe Armoured vehicles

Titel:	Finding the Koningsberg
Tydskrif:	Bulletin - Simon's Town Historical Society (1988)15:2, p.62-63 July.
Ekserp:	Describes the use of a seaplane as a reconnaissance aircraft to locate the German ship Koningsberg during World War 1
Trefwoorde:	Curtiss flying-boats Cutler, H.D. reconnaissance aircraft World War 1 naval warfare war Koningsberg battleships

Voorbeelde van resultate uit die titelveld:

Titel:	The South African white general election of 1987: shifting deckchairs or burning boats?
Tydskrif:	International affairs bulletin (1987)11:2, p.19-43
Outeur:	Lourens E & Kotze H
Ekserp:	Analyses voting patterns during the 1987 general election with particular reference to growing support to the Conservative Party. Speculates about future developments in electoral policies. Concludes that South African parliamentarians may well be on its last legs unless the white electorate can commit itself to further reform.

Titel:	Of passion, and leaking boats
Tydskrif:	Productivity SA(1993)19:8, p.5-6, Nov-Dec
Outeur:	Riley, A D
Ekserp:	Debates the appropriateness of leadership training methods. Briefly focuses on Clare Graves's value systems. Describes the nature of leadership at each value station in order to illustrate what kind of leadership behaviours different people would respond to. Highlights four basic themes for leadership training, each with its own assumptions.

IN U LIG

Soekterm: *Doop*

Voorbeelde van resultate uit die onderwerpsveld:

Titel: Bent u een ingewijde?
Outeur: Ouweel, W.J.
Tydskrif: Bijbel en wetenskap, 13(115):235-238, Okt.
Trefwoorde: Ingewydes
 Doop
 Sakramente

Titel: Peetouers
Tydskrif: (In Venster op die gesin : openhartige gesprekke met ouers, kinders en familie, Potchefstroom : IRS, F3, no.19 : Versamelwerke, no.19)
Trefwoorde: Peetouers
 Doop

Voorbeeld van resultate uit die titelveld:

Titel: Tradisie leef voort (Dr. J A Schutte doop eerste en derde geslag)
Outeur: Anoniem
Tydskrif: Kerkblad (GKSA), 11 Des.17
Trefwoorde: Verbondslyn
 Schutte, J A

Titel: Die doop van Rus
Outeur: Duvenage Johanna
Tydskrif: Kerkblad (GKSA), 29. Apr.6
Trefwoorde: Rusland

5.6 Analise van die resultate

5.6.1 Herwinning

Die resultate is hier ontleed bloot op grond van die hoeveelheid verwysings wat in elke soektog herwin is. Hiervolgens kan vasgestel word watter tipe term die beter resultate gelewer het.

In die 48 soektogte wat uitgevoer is (twee met elke indekseerterm: een in die onderwerpsveld en een in die titelveld), het die soektogte onder 'n saakhoof of deskriptor in alle gevalle ewe veel of meer verwysings herwin as wat die geval was met terme in die titel. Al is daar sover moontlik enkelwoord-indekseerterme gebruik, kan die aard van die terme (naamlik enkelvoudig teenoor saamgesteld) wel 'n invloed op die resultate hê. Titels is dikwels nie ekspressief ten opsigte van onderwerpsinhoud nie en bevat baiekeer ook nie vakterme nie. Dat daar by sulke enkelvoudige terme soos boats, abortion en aborsie beduidende verskille was, is eger insiggewend. Dit is woorde wat maklik in die titel van 'n dokument sou kon voorkom waarin oor sulke onderwerpe geskryf word.

5.6.2 Relevansie (presiesheid)

Om die onderwerpsinhoud van die dokument vas te stel, is die titel van die dokument sowel as die indekseerterme (en ekserpte waar daar is) wat in die bibliografiese rekord aangedui word, ontleed. Daarna is dit getoets vir relevansie op grond van die soekterm: eerstens waar die soekterm in die onderwerpsveld gebruik is (hier sal dit 'n toegekende indekseerterm wees) en tweedens waar die soekterm in die titelveld gebruik is (hier sal dit 'n voorbeeld van 'n afgeleide term wees).

Vervolgens sal 'n paar van die voorbeelde wat onder 5.5.2 gegee is, ontleed word.

5.6.2.1 Intydse katalogus

Onder die soekterm *academic achievement* is daar slegs een titel gevind wat naastenby hierdie woorde bevat, teenoor drie verwysings waar die woorde as toegekende term voorkom. Die titels wat dus aan dokumente gegee is wat oor die onderwerp handel, is nie verteenwoordigend van die inhoud nie. By die soekterm *mongolisme* is daar nie een titel waarin dit voorkom nie, teenoor drie rekords waarin dit in die onderwerpsveld voorkom. *Mongolisme* het dus geen dokumente herwin waar dit as afgeleide term gebruik is nie.

5.6.2.2 ERIC

Onder die soekterm *planetarium* (wat in Engelse en Afrikaanse titels dieselfde gespel sal wees) kom die woord net in een van die voorbeelde se titels voor, teenoor twee keer in die onderwerpsveld.

5.6.2.3 Repertorium

Hier is 'n voorbeeld waar 'n enkelvoudige woord soos *boats* (wat hoogs waarskynlik in titels sou kan voorkom), in net twee voorbeelde se titels voorkom, teenoor vier keer in die onderwerpsveld. Boonop het *boats* in die titels ook 'n nieletterlike betekenis.

5.6.2.4 In U Lig

Hier kom die soekterm *doop* in twee titels voor teenoor vier keer in die onderwerpsveld.

Samevatting

Uit hierdie ontleding van die soekresultate is dit duidelik dat daar 'n beduidende verskil tussen die voorkoms van terme in die titelveld en in die

onderwerpsveld is. Dit is duidelik dat die titel in die meeste gevalle nie 'n goeie aanduiding van die onderwerpsinhoud van die dokument is nie en dat daar deurgaans swakker soekresultate gelewer is as waar in die onderwerpsveld onder deskriptore of saakhoofde gesoek is.

HOOFSTUK 6

AFLEIDINGS EN GEVOLGTREKKING

6.1 Inleiding

Die doel van inligtingsontsluiting is om daardie inligting later te kan herwin. Inligting wat nie herwin kan word nie, is nie bruikbaar nie. Inligting in dokumente kan net van nut wees as gebruikers weet van die bestaan daarvan en ook waar dit gevind kan word. Die onderwerpsinhoud van dokumente moet dus pertinent onder die aandag van gebruikers gebring word. Dit vind in die IOHS deur middel van die indekseertaal plaas. Indekseertaal is soos reeds in 1.1 genoem, 'n tussentaal wat gebruik word om kommunikasie tussen die gebruikers van inligting en die inhoud van inligtingsbronne te bewerkstellig. Hierdie kommunikasie kan nie doeltreffend wees as die indekseertaal vaag of gebrekkig is nie.

Die doel van hierdie studie was om vas te stel of daar 'n verskil is in die kwaliteit van indekseerterme wat enersyds met toegekende indeksering en andersyds met afgeleide indeksering verkry is. Verder is beoog om vas te stel of menslike inligtingsverwerking enige rol daarin speel. Toegekende indeksering en afgeleide indeksering is met mekaar vergelyk, sodat vasgestel kon word watter tipe indekseermetode die beste kwaliteit soekterme gee en gevolglik ook beter ten opsigte van herwinning en presiesheid presteer.

6.2 Afleidings wat gemaak kan word

In 5.5.1 is gesien dat die gebruik van toegekende indekseerterme in 'n soektog deurgaans 'n hoër herwinning gee as terme wat in die titelveld gesoek is. Verder is gesien dat as die titels alleen ontleed word (vergelyk 5.5.2), die onderwerpsinhoud van die dokument meestal nie duidelik na vore kom nie. Daar is ook voorbeelde waar nierelevante dokumente herwin is deur net op een woord in die titel te konsentreer (byvoorbeeld in die soektog onder die soekterm *Viking*).

Alhoewel die resultate nie as absoluut aanvaar moet word nie, omdat die omvang van die ondersoek te beperk was, kan die afleiding tog gemaak word dat toegekende indekseerterme beter ten opsigte van herwinning sowel as presiesheid presteer. Die rede hiervoor moet gesoek word in die aard van die indekseerterme: by afgeleide terme is bloot woorde in die titel as soekpunte gebruik, terwyl 'n proses van menslike inligtingsverwerking vooraf by toegekende terme plaasgevind het, naamlik voordat moontlike indekseerterme vasgestel is.

In 3.4 is die Van Dijk-Kintsch teorie gebruik om aan te toon wat in die menslike geheue tydens die lees en verstaan van 'n dokument plaasvind. Hierdie teorie wys daarop hoe belangrik dit is om die semantiese struktuur van 'n teks te verstaan om 'n begrip daarvan te kan kry en dat slegs die mens in staat is om dit te kan doen. 'n Taal kan ook net produktief verstaan word as die sintaktiese en semantiese aspekte daarvan verstaan word. Hierdie kennis van taal word deur die mens aangeleer vanuit sy biologiese ontwikkeling, sy geskiedenis en die persoonlike ervaringe wat hy beleef. Omdat al hierdie prosesse nie deur 'n rekenaar nageboots kan word nie, is hulle nie in afgeleide indeksering teenwoordig nie.

Toegekende indeksering vind op dieselfde wyse plaas as waarvolgens die individu inligting in sy geheue verwerk. Die stappe wat plaasvind, vertoon verskeie ooreenkomste met die soekstrategie wat tydens die navraagproses gevolg word. Tydens toegekende indeksering word die onderwerp van die dokument vasgestel deur die verstandsrêels wat in die menslike geheue plaasvind, na te volg. Die "aboutness" of die *waaroor* van 'n dokument word geïdentifiseer deur die konsepte in 'n raamwerk te berg, sodat 'n kernagtige opsomming van die teks gemaak kan word. Hierdie stap is 'n belangrike vertrekpunt in die indekseerproses, omdat die inhoud van 'n dokument later in die soekproses akkuraat deur konsepte verteenwoordig moet word, sodat die inhoud vir 'n gebruiker kernagtig weergegee kan word.

Daar is egter tot dusver geen spesifieke en duidelike reëls oor hoe om te indekseer nie. Indekseerders ontleed en interpreteer dokumente volgens verstandsrêels. By afgeleide indeksering word die volledige teks van die dokument, of die belangrikste dele daarvan, gebruik om indeksering deur middel van 'n instruksie in 'n rekenaarprogram te bewerkstellig. Daar word

optimisties beweer dat 'n rekenaar die denkproses van die mens kan naboots en die inherente beperkinge van toegekende indeksering kan oorkom, maar soos wat reeds aangetoon is, is dit nie die geval nie.

6.3 Die rol van 'n beheerde woordeskat

'n Beheerde woordeskat is die enigste manier waarvolgens 'n onderwerp terminologies spesifiek beheer kan word om verwarring en vaagheid van benaminge uit te skakel. Geen rekenaarprogram is al in staat om die inhoud van 'n dokument algoritmies so te interpreteer dat die verwantskappe tussen terme aangedui kan word en elke vakgebied se terme gekies word volgens die vereistes wat gestel word nie. Slegs die menslike brein is in staat om die verwantskappe tussen terme op grond van logiese beredenering en kennis te definieer. Laasgenoemde word verkry uit die proses van inligtingverwerking en uit ervaring wat in 'n raamwerk van verwante kennis gestruktureer word. Al hierdie kennis word gebruik om 'n inligtingsbehoefte te identifiseer en te verwoord. Afgeleide indeksering is tot dusver - met al die voordele van kunsmatige intelligensie - nog nie in staat om self 'n gebruiker se soektog te analiseer en moontlike verwante terme voor te stel nie. 'n Gekontroleerde woordeskat verskaf aan die gebruiker 'n lys woorde waarmee hy die konsepte in sy geheue in ooreenstemming kan bring.

In sommige van die vakdissiplines, soos chemie, wiskunde, fisika en ingenieurswese, is die terme wat in publikasies en die spreektaal gebruik word, oor die algemeen eie aan die vak en beskrywend presies. In die geesteswetenskaplike dissiplines is dit baiekeer nie die geval nie. Vakterme is minder geïk en terme uit verskillende dissiplines oorvleuel dikwels. Dieselfde konsep kan ook deur verskillende terme uitgedruk word, met die gevolg dat inligting oor die hele inligtingsspektrum versprei word. Soektogte kan gekompliseer word, aangesien gebruikers deur massas inligting moet soek en self woordeskatbeheer moet toepas sonder die aanduiding van alternatiewe soekterme. Die gebruiker is dus aan sy eie diskresie oorgelaat. 'n Beheerde woordeskat, waar variante vorme beheer word, kan die probleme oplos.

'n Beheerde woordeskat kan egter ook die prestasie van 'n IOHS negatief beïnvloed. Die grootste remskoen wat veral tydens herwinning na vore kom, is dat die indekseerder nie akkuraat genoeg geïndekseer het nie. Dit veroorsaak dat die terme nutteloos binne die IOHS is. Dokumente word herwin wat nierelevant is. Die woordeskat waartoe die gebruiker toegang verkry is nie presies nie en belangrike terme sal misgekyk of verkeerd geïnterpreteer word.

6.4 Gevolgtrekking

Uit die eksperiment wat uitgevoer is, het dit duidelik geblyk dat toegekende terme beter as afgeleide terme presteer het. Die gevolgtrekking wat hieruit gemaak kan word, is dat 'n indekseerder met sy sosiale oorsprong, kennis van taal en menslike oordeel die beste toegerus is om die proses van indeksering uit te voer. Die fisiese inligtingsverwerkingsproses wat plaasvind wanneer die teks van 'n dokument deur middel van taalreëls verwerk, interpreteer en kernagtig deur 'n paar indekseerterme saamgevat word, is die sleutel tot die prestasievermoë van 'n IOHS. 'n Rekenaarprogram is nie in staat om taal op dieselfde wyse as die mens te verwerk nie. Die mens kán dit vanuit 'n menslike milieu verwerk en orden, want taal word *deur* die mens *vir* die mens geskep. Die mens kan dus indekseerterme toeken wat die onderwerpsinhoud van 'n dokument korrek weergee. Woorde wat slegs uit die titel of teks afgelei is, is nie altyd 'n goeie aanduiding van die intellektuele onderwerpsinhoud nie en lewer dus swakker kwaliteit terme. Menslike inligtingsverwerking speel 'n belangrike rol in die kwaliteit van indekseerterme. Milstead (1995:93) som dit soos volg op:

"The best statistical and syntactic analysis cannot substitute for access based on the actual intellectual content of a document."

(Met "statistical and syntactic analysis" word na 'n rekenaarproses verwys.)

6.5 Slotopmerking

Wessel (1975:13) sê dat n t die mens oor die kreatiwiteit beskik om 'n onderskeid te tref tussen die betekenis van *woorde* en *woorde* wat gebruik word om na ander *woorde* te verwys, asook om *woorde* te gebruik om die nieverbale betekenis van *woorde* uit te druk. Hierdie kreatiwiteit is 'n komplekse begrip as sodanig. Dit is 'n aanleg wat deur God geskep is en die mens gebruik dit te alle tye om die aandag toe te spits op dit waaraan hy 'n behoefte het. Kreatiwiteit word gekenmerk deur *oorspronklikheid*, *herkenbaarheid* en *doelgerigtheid* wat in alle handeling wat deur die mens uitgevoer word, na vore kom. Daarom sal menslike oordeel selfs tydens rekenaarmatige indeksering bly geld om ontsluiting en herwinning te bewerkstellig.

BIBLIOGRAFIE

AITCHISON, J. & GILCHRIST, A. 1987. Thesaurus construction: a practical manual. London : Aslib. 173 p.

ALBERICO, R. & MICCO, M. 1990. Expert systems for reference and information retrieval. Westport : Mecker. 395 p. (Supplements to computers in libraries; no. 10.)

ALURI, R., KEMP, D.A. & BOLL, J.J. 1991. Subject analysis in online catalogs. Englewood, Colo : Libraries Unlimited. 303 p.

ANDERSON, J.D. 1985. Indexing systems: extensions of the mind's organizing power. *Information and behavior*, (1):287-323.

ANDRE, T. 1979. Does answering high-level questions while reading facilitate productive learning? *Review of educational research*, 49(2):280-318.

ARMSTRONG, C. J. & LARGE, J.A. 1988. Developing search strategies in manual of online search strategies. Aldershot : Gower. 43 p.

BARRY, C.L. 1994. User-defined relevance criteria: an exploratory study. *Journal of the American Society for Information Science*, 45(3):149-159, Apr.

BDI

kyk

VAN SWIGCHEN, P.J. & SLOT, E.J.

BEGHTOL, C. 1986. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of documentation*, 42(2):84-113, Jun.

BLAIR, D.C. & MARON, M.E. 1985. An evaluation of retrieval effectiveness for a full-text document retrieval system. *Communications of the ACM*, 28(3):289-299, Mar.

BOON, J.A. 1979. Enkele perspektiewe in die ontsluiting van inligtingsbronne. Johannesburg. 24 p. (Inougurele rede - RAU.)

BOON, J.A. 1982. User evaluation of information retrieval systems: some methodological considerations. *Mousaion II*, 11:1-96.

BOON, J.A. 1976. Kriteria en meeteenhede by die evaluering van gerekenariseerde SDI en KSI-dienste. *Suid-Afrikaanse biblioteke*, 43(3):83-88, Jan.

BOURNE, C.P. 1963. *Methods of information handling*. New York : John Wiley & Sons. 241 p.

BUSHA, C.H., & HARTER, S.P. 1980. *Research methods in librarianship: techniques and interpretation*. New York, N.Y. : Academic Press. 417 p.

CHU, C.M. & O'BRIEN, A. 1993. Subject analysis: the critical first stage in indexing. *Journal of information science*, 19(6):439-454.

CLEVELAND, D.B. & CLEVELAND, A.D. 1983. *Introduction to indexing and abstracting*. Littleton, Colo. : Libraries Unlimited. 209 p.

CLEVELAND, D.B. & CLEVELAND, A.D. 1990. *Introduction to indexing and abstracting*. Littleton, Colo. : Libraries Unlimited. 329 p.

CRAIK, F.I.M. 1979. Human memory. *Annual review of psychology*, 30:63-102.

CRAIK, F.I.M. & LOCKHART, R.S. 1972. Levels of processing: a framework for memory research. *Journal of verbal learning and verbal behavior*, 11:671-684.

DE WET, J.L., MONTEITH, J.L. de K. & VAN DER WESTHUIZEN, G.J. 1981. Opvoedende leer. Durban : Butterworths. 470 p.

DILLON, M & MCDONALD, L.K. 1983. Fully automatic book indexing. *The journal of documentation*, 39(3):135-154, Sep.

DILLON, M. & GRAY, A.S. 1983. FASIT: a fully automatic syntactically based indexing system. *Journal of the American Society for Information Science*, 34(2):99-108, Mar.

FARRADANE J. 1979. The nature of information. *Journal of information science*, 1:13-17.

FARRADANE, J. 1980. Knowledge, information, and information science. *Journal of information science*, 2:75-80.

FARROW, J. 1991. A cognitive process model of document indexing. *Journal of documentation*, 47(2):149-166, Jun.

FARROW, J. 1994. Indexing as a cognitive process. (*In Encyclopedia of information science*, 53. Supplement, 16:155-171.)

FEINBERG. H. 1973. Title derivative indexing techniques: a comparative study. Metuchen, N.J. : Scarecrow Press. 297 p.

FOSKETT, A.C. 1986. The subject approach to information. 4th ed. London : Bingley. 574 p.

FOURIE, I. 1990. Inligtingkunde (INL). Pretoria : Universiteit van Suid-Afrika. 184 p. (Studiegids ; INL302-6.)

FROHMANN, B. 1990. Rules of indexing: a critique of mentalism in information retrieval theory. *The journal of documentation*, 46(2):81-101, Jun.

FROST, C.O. 1989. Title words as entry vocabulary to LCSH: correlation between assigned LCSH terms and derived terms from titles in bibliographic records with implications for subject access in online catalogs. *Cataloging and classification quarterly*, 10(1/2):165-179.

GLASS, A.R., HOLYOAK, K.J. & SANTA, J.L. 1979. *Cognition*. Reading, Mass. : Addison-Wesley. 521 p.

GOUWS, L.A., LOUW, D.A., MEYER, W.F. & PLUG, C. 1979. *Psilogiewoordeboek*. Johannesburg : McGraw-Hill. 484 p.

HARTER, 1986. *Online information retrieval: concepts and techniques*. Orlando : Academic Press. 259 p.

HARTLEY, R.J., ed. 1990. *Online searching principles and practice*. London : Bowker-Saur. 387 p.

HUTCHINS, W.J. 1978. The concept of "aboutness" in subject indexing. *Aslib proceedings*, 30(5):172-181, May.

ISO STANDARD

kyk

INTERNATIONAL ORGANISATION FOR STANDARDISATION.

INTERNATIONAL ORGANISATION FOR STANDARDISATION. 1985. *Documentation: methods for examining documents, determining their subjects, and selecting indexing terms.* (ISO Standard ; 5963-1985(E).)

KLIX, F. & HOFFMANN, J. 1980. *Cognition and memory*. Amsterdam : North-Holland. (Advances in psychology ; no.5.) 306 p.

KUHLTHAU, C.C. 1991. Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5):361-371, Jun.

KUHLTHAU, C.C. 1994. Students and the information search process: zones of intervention for librarians. (*In* Godden, I.P. *Advances in librarianship*, 18. San Diego, Calif : Academic Press. p. 57-72)

LANCASTER, F.W. 1968. *Information retrieval systems: characteristics, testing, evaluation.* New York, NY : Wiley. 222 p.

LANCASTER, F.W. 1979. *Information retrieval systems: characteristics, testing and evaluation.* 2nd ed. New York, N.Y. : Wiley. 381 p.

LANCASTER, F.W. 1986. *Vocabulary control for information retrieval.* 2nd ed. Arlington, VA : Information Resources Press. 207 p.

LANCASTER, F.W. 1991. *Indexing and abstracting in theory and practice.* London : Library Association. 328 p.

LANGRIDGE, D.W. 1989. *Subject analysis: principles and procedures.* London : Bowker-Saur. 146 p.

LEONARD, L.E. 1977. *Inter-indexer consistency studies, 1954-1975: a review of the literature and summary of study results.* Urbana, Il. : University of Illinois Graduate School of Library Science. 51 p. (Occasional papers; no. 131.)

LOUW, D.A. 1986. *Inleiding tot die psigologie.* Johannesburg : McGraw-Hill. 649 p.

MILSTEAD, J.L. 1984. *Subject access systems: alternatives in design.* Orlando : Academic Press. 212 p.

MILSTEAD, J.L. 1995. *Invisible thesauri: the year 2000, semantic access to full text.* *Online & CD Rom review*, 19(2):93-94, Apr.

PAO, M.L. 1989. *Concepts of information retrieval.* Englewood, Colo : Libraries Unlimited. 285 p.

- PARK, T.K.** 1994. Toward a theory of user-based relevance: a call for a new paradigm of inquiry. *Journal of the American Society for Information Science*, 45(3):135-141, Apr.
- REED, S.K.** 1992. Cognition: theory and applications. 2nd ed. Pacific Grove, Calif. : Brooks/Cole. 348 p.
- REES, A.M. & SARASEVIC, T.** 1966. The measurability of relevance. *Proceedings of the American Documentation Institute*, 3:225-234.
- ROWLAND, G.T. & McGUIRE, J.A.** 1971. The mind of man: some views and a theory of cognitive development. New Jersey, NY : Prentice-Hall. 167 p.
- ROWLEY, J.E.** 1987. Organising knowledge: an introduction to information retrieval. Aldershot : Gower. 454 p.
- ROWLEY, J.E.** 1988. Abstacting and indexing. London : Bingley. 181 p.
- SCHNEIDER, W. & SHIFFRIN, R.M.** 1977. Controlled and automatic human information process : detection, search and attention. *Psychological review*, 84(1):1-53.
- SOERGEL, D.** 1985. Organizing information: principles of data base and retrieval systems. San Diego, Calif. : Academic Press. 450 p.
- SOERGEL, D.** 1994. Indexing and retrieval performance: the logical evidence. *Journal of the American Society for Information Science*, 45:589-599, Sep.
- SVENONIUS, E.** 1986. Unanswered questions in the design of controlled vocabularies. *Journal of the American Society for Information Science*, 37(5):331-340, Sep.

- TAYLOR, R.S.** 1968. Question negotiation and information seeking in libraries. *College and research libraries*, 29(3):178-194.
- TINKER, J.F.** 1966. Imprecision in indexing. *American documentation*, 17:93-102.
- TODD, R.J.** 1992. Academic indexing: what's it all about? *The indexer*, 18(2):101-104, Oct.
- UNESCO.** 1975. UNISIST indexing principles. Paris : UNESCO. 10 p.
- VAN DIJK, T.A. & KINTSCH, W.** 1983. Strategies of discourse comprehension. San Diego, Calif. : Academic Press. 418 p.
- VAN SWIGCHEM, P.J. & SLOT, P.J.** *reds.* 1990. BDI-terminologie : het verklarend woordenboek van Nederlandse termen op het gebied van bibliotheek en documentaire informatie met vertalingen in het Engels, Frans, Duits, Spaans. Den Haag : Nederlands Bibliotheek en Lektuur Centrum. 493 p.
- VICKERY, B.C.** 1968. Analysis of information. (*In Kent, A. & Lancour, H. ed. Encyclopedia of library and information science*, 1:355-384.)
- VICKERY, B.C. & VICKERY, A.** 1987. Information Science in theory and practice. London : Butterworths. 384 p.
- WELLISCH, H.H.** 1989. The literature of indexing. (*In Weinberg, B.H., ed. Indexing the state of our knowledge and the state of our ignorance.* Medford : Learned Information. p. 1-14.)
- WESSEL, A.E.** 1975. Computer-aided information retrieval. Los Angeles : Melville. 176 p. (Information science series.)