



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Relevance judgements in information retrieval

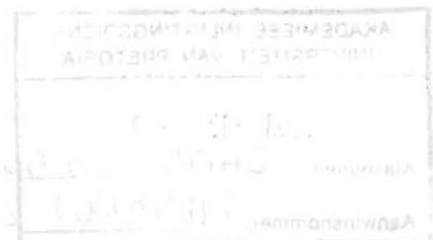
by

Erica Cosijn

Submitted in fulfilment of part of the requirements
for the degree of
Doctor of Information Science
In the Faculty of Engineering, Built Environment and
Information Technology
University of Pretoria

Supervisor: Prof T J D Bothma

May 2003





ACKNOWLEDGEMENTS

Many people have contributed and supported me during the years of study of which this thesis is the culmination, and I would like to thank them:

- My supervisor, Professor Theo Bothma, as well as Professors Peter Ingwersen and Kalervo Järvelin for encouragement, support and advice.
- The originators of the DISSANet programme, Professors Irene Wormell, Peter Ingwersen, Theo Bothma and Rocky Ralebipi. This three year programme to facilitate the development of IS research in South Africa was funded by Danida.
- The researchers and students at the doctoral workshops in Borås (August 2000) and Copenhagen (November 2000) for helpful discussions and advice.
- Susanna Oosthuizen and Professor Niels-Ole Pors for help on the questionnaire construction, and Rina Owen and Dr Hermi Boraine for the analysis of the empirical data.
- The volunteers who took part in the empirical study.
- My family for encouraging me and believing in me.

ABSTRACT

Recent studies in the measurement of relevance criteria across stages of document evaluation concludes that the findings "...suggest a need for continued work to map or array relevance criteria across information search process stages, variations in document representations, tasks and contexts." (Tang & Solomon, 2001).

This thesis aims to develop such a model. In this study the attributes and manifestations of relevance as defined by Saracevic (1996) are modelled in a matrix in order to define the various relevance types more clearly. From this modelling process an array of relevance types are derived, namely algorithmic or systems relevance, topical relevance, cognitive relevance, situational relevance, socio-cognitive relevance and affective relevance.

These identified relevance types are then modelled on an existing cognitive model of information transfer, as defined by Ingwersen (1996). The Ingwersen model was utilized because it explores the multifunctional and cognitive array of representations of both the information objects, and the cognitive space of the user, both as influenced by the environment. By the re-organization of this model, the different types of relevance were shown to operate in different dimensions (and over time) of the information retrieval process. This model has the added advantage that it may be possible to specify whether relevance judgments were made during the work task or the search task execution.

The viability of this model is then indicated by utilizing the published results of two empirical studies, namely that of Barry and Schamber 1998) and Vakkari and Hakala (2000). The criteria identified in this process were then consolidated, analysed and allocated to the corresponding manifestations of relevance and relevance types as identified and modelled, excluding the more "objective" relevance types, over which the user does not have much control.

The research questions posed relates to various aspects of the model, such as the relationship between affective relevance and the other subjective relevance types, the existence of socio-cognitive relevance, the relationship between cognitive and socio-cognitive relevance and the judgements of documents within work task domains. The model was then tested, both in terms of the validity of the construct and the research questions stated. The empirical testing was done by means of questionnaires, once the work task of the respondent has been completed.

The structure of the thesis is as follows: Chapter 2 is a literature review tracing the history of relevance research as well as the multidimensional and dynamic nature and the interdisciplinary research involved. Through this overview it becomes clear that there is a need to model relevance types in terms of a more holistic approach, and therefore the development of such a model has been formulated in Chapter 3. Chapter 4 describes the construction of the questionnaires in order to test the model developed in Chapter 3. The results gathered by means of the questionnaires are presented and discussed in Chapter 5. The conclusion and discussion of the results in terms of the model developed are documented in Chapter 6. Chapter 7 reviews the larger significance of the results in terms of possible practical implementation of the findings.



CHAPTER 1: INTRODUCTION	1
1.1. The concept of relevance within the context of Information Science	1
1.2. The research problem	2
1.3. Aims, goals and research objectives - research questions	2
1.4. Methodology	3
1.5. Outline of thesis	4
CHAPTER 2: LITERATURE REVIEW	5
2.1. Introduction: demarcation of literature covered	5
2.2. Relevance defined	5
2.2.1. Definitions and conceptions of relevance	6
2.2.2. The history of relevance	9
2.2.2.1. Before 1958	10
2.2.2.2. 1959 – 1976	11
2.2.2.3. 1977 onwards	11
2.3. The multidimensionality and dynamic nature of relevance	12
2.3.1. The interdisciplinary nature of relevance	13
2.3.2. Degrees of relevance	14
2.3.3. The dynamic nature of relevance	15
2.4. Research with implications for relevance	16
2.5. Summary of main conclusions based on literature review	16
CHAPTER 3: DEVELOPING A THEORETICAL FRAMEWORK	18
3.1. Attributes of relevance	19



3.2. Manifestations of relevance	20
3.3. Attributes and manifestations of relevance: What are the connections?	22
3.3.1. Relation	23
3.3.2. Intention	26
3.3.3. Context	28
3.3.4. Inference	28
3.3.5. Interaction	29
3.3.6. Motivational relevance as intentionality	31
3.4. The modified relevance model	32
3.5. Some consequences of relevance variety	35
3.6. Relevance types	36
3.6.1. Algorithmic relevance	37
3.6.2. Topicality	38
3.6.3. Cognitive relevance or pertinence	38
3.6.4. Situational relevance	39
3.6.5. Socio-cognitive relevance	40
3.6.6. Affective relevance	40
3.7. The contexts of relevance judgements in the information seeking process	41
3.7.1. Social/organizational domain	48
3.7.2. Defining/perceiving the work task	48
3.7.3. The individual's cognitive space	49
3.7.4. Statement of information need	49
3.7.5. Request and request formulation	49
3.7.6. Interface/Intermediary	50
3.7.7. Information objects	50
3.8. Work task and search task as depicted in the model	51
3.9. User criteria for relevance judgments	53



3.10. Summary and conclusions	59
CHAPTER 4: RESEARCH DESIGN AND METHODOLOGY	61
4.1. Defining the research question	61
4.2. Construction of the questionnaire	63
4.2.1. Length of the questionnaire	64
4.2.2. Language and vocabulary	64
4.2.3. Wording of the questions	65
4.2.4. Sequence of the questions	65
4.2.5. Types of questions	66
4.2.6. Question content and selection	66
4.3. The questionnaire	69
4.3.1. Section A: context of information seeking and use	75
4.3.2. Section B: documents used to complete the work task	75
4.3.3. Section C: documents retrieved and read, but not used	84
4.4. Sample design and sampling methods	90
4.5. Data collection methods	92
4.6. Data capturing and data editing	93
4.7. Data analysis	93
4.8. Limitations of the methodology	93
4.9. Summary	94
CHAPTER 5: PRESENTATION AND DISCUSSION OF RESULTS	95
5.1. Sample profiles	95
5.2. Research questions: Data analysis and results	97
5.2.1. The main research question	97
5.2.2. Sub-question 1	101



5.2.3. Sub-question 2	102
5.2.4. Sub-question 3	115
5.2.5. Sub-question 4	117
5.2.6. Sub-question 5	122
5.2.7. Sub-question 6	127
5.3. Summary of findings	136
CHAPTER 6: CONCLUSIONS REGARDING MODELLING AND EMPIRICAL DATA	138
6.1. Conclusions from the literature review	138
6.2. Conclusions regarding the model developed	138
6.3. Conclusions for future research	139
CHAPTER 7: IMPLICATIONS OF THE MODEL FOR IR RESEARCH	140
7.1. Algorithmic relevance	141
7.2. Topicality	142
7.2.1. High order knowledge representation	143
7.2.2. Fuzzy and parallel IR	144
7.3. Cognitive relevance / pertinence	144
7.4. Situational relevance	147
7.5. Socio-cognitive relevance	148
7.6. Affective relevance	151
7.7. Conclusions	151



APPENDIX A: QUESTIONNAIRES	163
Section A:	164
Section B:	167
Section C:	171
APPENDIX B: COVER LETTER & LETTER OF INFORMED CONSENT	175
APPENDIX C: DEGREES OF RELEVANCE WITHIN WORK TASK	178
APPENDIX D: RELEVANCE TYPES BY WORK TASK	180
APPENDIX E: RELEVANCE JUDGEMENTS BY TASK TYPE	183
APPENDIX F: THE IMPACT OF SOCIO-COGNITIVE RELEVANCE CONSIDERATIONS ON DOCUMENT USE	185



LIST OF FIGURES

Figure 3.1.	A classification of matching methods	37
Figure 3.2.	Cognitive model of information transfer	45
Figure 3.3.	Relevance types, work task and search task in information seeking and retrieval	47
Figure 5.1.	Reasons for use by work task	106
Figure D.1.	Type of relevance judgements by work task	182

LIST OF TABLES

Table 3.1.	Attributes of relevance	20
Table 3.2.	Manifestations of relevance	21
Table 3.3.	Attributes and manifestations of relevance	22
Table 3.4.	Revised table of relevance types and attributes	33
Table 3.5.	Barry and Schamber relevance criteria	54
Table 3.6.	Vakkari and Hakala relevance criteria and subcategories	55
Table 3.7.	Criteria pertaining to relevance types	58
Table 4.1.	Variables to questions matrix: Section B	67
Table 4.2.	Variables to questions matrix: Section C	68
Table 4.3.	Codes used for relevance types in questionnaires	70
Table 4.4.	Relevance criteria within relevance types	72
Table 4.5.	Questionnaires completed	91
Table 5.1.	Summary of survey sample	96
Table 5.2.	Analysis of topical relevance judgements together with other subjective relevance types in terms of reasons for use – Question 2	100
Table 5.3.	Analysis of topical relevance judgements together with other subjective relevance types in terms of reasons for use – Question 5	101
Table 5.4.	Reasons for use/usefulness of document by work task	105
Table 5.5.	Relevance judgements by task type - documents cited	117
Table 5.6.	Nesting of relevance types by number of occurrences	120
Table 5.7.	Visualization of the nesting characteristics of relevance types according to empirical data	120
Table 5.8.	Negative affective relevance judgements made together with reasons for document use	124
Table 5.9.	Negative affective relevance judgements made together with reasons for regarding a document as useful	125
Table 5.10.	Negative affective relevance judgements made together with stated knowledge of the terminology of document	125
Table 5.11.	Negative affective relevance judgements made together	126



	with reasons for regarding a document as useful - not cited	
Table 5.12.	Negative affective relevance judgements made together with reasons for not using a document	126
Table 5.13.	Negative affective relevance judgements made together with relationship with the author of a document - not cited	127
Table 5.14.	Impact of socio-cognitive relevance considerations on document use	129
Table 5.15.	Impact of socio-cognitive relevance considerations on the non-use of documents	129
Table 5.16.	Cognitive relevance judgements made in conjunction with socio-cognitive relevance regarding academic standards of documents	132
Table 5.17.	Cognitive relevance judgements made in conjunction with socio-cognitive relevance regarding viewpoint congruence of documents	132
Table C.1.	Reasons for use/usefulness by work task and degree of usefulness	179
Table D.1.	Relevance types by work task	181
Table E.1.	Relevance judgements by task types – documents cited	184
Table F.1.	The impact of socio-cognitive relevance considerations on document use	186



"Relevance will serve its purpose, but will decline as the realization slowly comes that an individual's information need is so complex ... The gradually increasing awareness of a human's incapability of stating his true need in simple form will tend to pull the rug out from under many IR system evaluation studies which will have been done in the meanwhile." (Doyle, 1963)

"Our understanding of relevance in communication is so much better, clearer, deeper, broader than it was when information science started after the Second World War. But there is still a long, long way to go." (Saracevic, 1975)

"We consider the pursuit of a definition of relevance to be amongst the most exciting and central challenges of information science, one whose solution will carry us into the 21st century." (Schamber et al., 1990)

"Relevance is a necessary part of understanding human behaviour. The field should be encouraged by commonalities across perspectives, not discouraged by disagreements. Relevance presents a frustrating, provocative, rich, and - undeniably - relevant area of inquiry." (Schamber, 1994)

"Nobody has to explain to users of IR systems what relevance is ... People understand relevance intuitively." (Saracevic, 1996)