

10

**Internet Searching as a study  
aid for Information  
Technology and Information  
Systems learners at a tertiary  
level**

by Melius Weideman

**Thesis presented for the Degree of  
DOCTOR OF PHILOSOPHY  
in the Department of Information and  
Library Studies**

UNIVERSITY OF CAPE TOWN

October 2001

# DECLARATION

I, the undersigned, hereby declare that the work done towards this qualification has been my own work, except where stated otherwise. All references used were accurately recorded, and any opinions expressed herein are my own and not necessarily those of the University of Cape Town.

STUDENT:

Signature: Signed by candidate  
M. Weideman

Date: 4-12-2001

SUPERVISOR:

Signature: P.G. Underwood

Date: 4-12-2001

# ACKNOWLEDGEMENTS

I should like to thank the following, without whose help this project would not have been possible:

1. My Creator for giving me the staying power.
2. My wife Alta, children Tanya and Oelof, for their understanding, love and support all the way, and for often coping without a husband and father.
3. My colleague Corrie Strümpfer, for helping with the statistical analysis.
4. My ex-student Herman Reinders, for helping with the intricacies of Java.
5. All the controllers at the Institutions in South Africa, Slovenia and in the USA for their assistance and preparation for my experiments.
6. My colleague Prof van Aswegen and friend Clare le Roux for their meticulous proofreading of this document.
7. Prof Underwood for his guidance, patience, gentle manner and understanding.
8. The Cape Technikon for study leave and financial assistance, the National Research Foundation, Acer Africa and Hewlett-Packard SA for financial assistance.

# ABSTRACT

## **Internet Searching as a study aid for Information Technology and Information Systems learners at a tertiary level.**

In this thesis, the author attempted to develop a method to help Information Technology/Systems learners find relevant information on the Internet. The literature indicated that it is essential that learners should be able to retrieve relevant information from electronic sources. However, it was also stated repeatedly that searching on the Internet using standard search engines is not an easy task. It was also noted that a move was taking place away from traditional teaching methods to those with more learner involvement, making use of new computer and communication technologies.

Initial experiments were done with IT/IS learners to determine how and where they search on the Internet, and what degree of success they had. The most important data gathered from these experiments was the lack of search strategy displayed by learners; the search engines chosen by them; and their success rate. Only 32.2% of all learners in this study managed to find one piece of relevant information in 30 minutes without any assistance.

The data was used to design and later improve a tool to guide them in their searching endeavours. This tool, called "Finder Of Information On The Internet" (FOIOTI, at <http://www.mwe.co.za>), was then extensively tested by measuring searching success, with and without using it. The data was gathered by examining and summarizing the forms completed by the learners during the searching experiments.

During this study the author found that most learners had little or no training on Internet usage; often worked on the Internet; could not specify their search properly; used very few of the operators offered by search engines; and had a low success rate in finding relevant data.

The two final phases of experiments proved that FOIOTI, as a searching tool, was successful. During these two phases, 71.0% of the participants claimed that they found the specified information when using FOIOTI within 30 minutes.

A total of 1109 learners from three continents and 20 institutions took part in the study, spread over 46 sessions. Every session was administered personally by the author.

This research project has contributed to the existing body of knowledge on Information Retrieval and education by having developed a tool that enhances learner involvement in the learning process. It enables educators to explore easier alternatives of locating educational resources by drawing on the experience of other website authors in their field. It also entices the average learner to re-skill him/herself on the use of an exciting and promising new technology: Information Retrieval through the Internet.

# TABLE OF CONTENTS

## CHAPTER 1 – BACKGROUND AND PROBLEM STATEMENT

1.1	INTRODUCTION	1
1.2	THE LEARNING SPACE	2
	1.2.1 SCHOOLING	2
	1.2.2 INSTITUTIONS OF HIGHER EDUCATION	2
	1.2.3 THE LEARNERS	3
	1.2.4 EDUCATION IN SOUTH AFRICA: CURRENT SITUATION	4
1.3	MOTIVATION	5
1.4	LEARNER PARTICIPATION	6
1.5	RESOURCES	6
	1.5.1 USE OF RESOURCES	7
	1.5.1.1 Textbooks	8
	1.5.1.2 Handouts and transparencies	8
	1.5.1.3 The library	10
	1.5.2 THE INTERNET	10
1.6	SUMMARY	14
1.7	CONCLUSION	15

## CHAPTER 2 – LITERATURE REVIEW

2.1	INTRODUCTION	16
2.2	HISTORY AND STATUS QUO	17
	2.2.1 INFORMATION SCIENCE AND INFORMATION RETRIEVAL	17
	2.2.1.1 Introduction	17
	2.2.1.2 History of IR	17
	2.2.1.3 Technological background	21
	2.2.1.4 Content representation background	22
	2.2.1.5 Human perspective background	25
	2.2.1.6 Status Quo	26
	2.2.2 THE INTERNET AND HYPERLINKS	27
	2.2.2.1 History of the Internet	27
	2.2.2.2 History of Hyperlinks	35
	2.2.3 IT IN EDUCATION	36
	2.2.3.1 Educator versus Learner	37
	2.2.3.2 Educational methodology and technology	37
2.3	THE LIBRARY	40
	2.3.1 THE ROLE OF THE WWW	40
	2.3.2 THE WWW AS SOURCE OF EDUCATIONAL INFORMATION	42
2.4	CASE STUDIES AND EXPERIMENTS	44
	2.4.1 INFORMATION RETRIEVAL	44
	2.4.2 SEARCHER BEHAVIOUR	47
	2.4.2.1 Non-Internet searching	48
	2.4.2.2 Internet searching	50

<b>2.5</b>	<b>SEARCH ENGINES</b>	<b>51</b>
	2.5.1 INTRODUCTION	51
	2.5.2 INTERFACE DESIGN	54
	2.5.3 OPERATORS	55
	2.5.4 QUERY FORMULATION	57
	2.5.5 SEARCH ENGINE EVALUATION	59
	2.5.6 RELEVANCE, RECALL, PRECISION AND RANKING	61
	2.5.6.1 Introduction	61
	2.5.6.2 Relevance	62
	2.5.6.3 Recall	64
	2.5.6.4 Precision	66
	2.5.6.5 Ranking	67
	2.5.7 MULTIPLE INDICES	68
	2.5.8 SEARCHING SUCCESS RATE	69
<b>2.6</b>	<b>SEARCHING METHODOLOGIES</b>	<b>72</b>
<b>2.7</b>	<b>SUMMARY</b>	<b>77</b>

## CHAPTER 3 – LITERATURE ANALYSIS

<b>3.1</b>	<b>INTRODUCTION</b>	<b>78</b>
<b>3.2</b>	<b>HISTORY AND STATUS QUO</b>	<b>78</b>
	3.2.1 INFORMATION SCIENCE AND INFORMATION RETRIEVAL	78
	3.2.2 THE INTERNET AND HYPERLINKS	79
	3.2.3 IT IN EDUCATION	81
<b>3.3</b>	<b>THE LIBRARY</b>	<b>84</b>
	3.3.1 THE ROLE OF THE WWW	84
	3.3.2 THE WWW AS SOURCE OF EDUCATIONAL INFORMATION	85
<b>3.4</b>	<b>CASE STUDIES AND EXPERIMENTS</b>	<b>86</b>
	3.4.1 INFORMATION RETRIEVAL	86
	3.4.2 SEARCHER BEHAVIOUR	87
<b>3.5</b>	<b>SEARCH ENGINES</b>	<b>88</b>
	3.5.1 INTRODUCTION	88
	3.5.2 INTERFACE DESIGN	88
	3.5.3 OPERATORS	89
	3.5.4 QUERY FORMULATION	89
	3.5.5 SEARCH ENGINE EVALUATION	90
	3.5.6 RELEVANCE, RECALL, PRECISION AND RANKING	90
	3.5.7 MULTIPLE INDICES	92
	3.5.8 SEARCHING SUCCESS RATE	92
<b>3.6</b>	<b>SEARCHING METHODOLOGIES</b>	<b>92</b>
<b>3.7</b>	<b>SUMMARY</b>	<b>93</b>

## CHAPTER 4 – RESEARCH METHODOLOGY

<b>4.1</b>	<b>INTRODUCTION</b>	<b>94</b>
	4.1.1 OBJECTIVE	94
	4.1.2 HYPOTHESIS	95
	4.1.2.1 Use of Internet technologies	95
	4.1.2.2 The information seeking facilities of the Internet	95
	4.1.2.3 Results	95
	4.1.2.4 Acquisition of skills	95
	4.1.3 MODEL	96
	4.1.4 POPULATION	97
	4.1.5 SAMPLE	98
	4.1.6 CONTROL	100

<b>4.2</b>	<b>RESEARCH METHODS</b>	<b>100</b>
	4.2.1 POSTAL QUESTIONNAIRE	101
	4.2.1.1 Non-response	101
	4.2.1.2 Questionable validity	102
	4.2.1.3 Questionable reliability	102
	4.2.1.4 Respondent bias	102
	4.2.1.5 Control	102
	4.2.1.6 Conclusion	103
	4.2.2 INTERNET QUESTIONNAIRE	103
	4.2.3 CASE STUDY	103
	4.2.4 PERSONAL INTERVIEW	104
	4.2.5 TELEPHONIC INTERVIEW	104
	4.2.6 LABORATORY EXPERIMENT	105
	4.2.7 FIELD EXPERIMENT	105
<b>4.3</b>	<b>PHASE 1 – DESIGN AND TEST INSTRUMENT: FORM 1</b>	<b>107</b>
<b>4.4</b>	<b>PHASE 2 – INTERPRET RESULTS, MODIFY FORM 1 &gt; FORM 2</b>	<b>109</b>
	4.4.1 FOCUS	109
	4.4.2 FEEL	109
	4.4.3 THREE SEARCHES	109
	4.4.4 FEEDBACK	109
	4.4.5 FALSE YES	109
	4.4.6 SUCCESS RATE	110
	4.4.7 HEADING	111
	4.4.8 READABILITY	111
	4.4.9 THREE SEARCHES	111
	4.4.10 FEEDBACK	111
	4.4.11 IDENTIFICATION	111
<b>4.5</b>	<b>PHASE 3 – TEST FORM 2 AS INSTRUMENT</b>	<b>111</b>
<b>4.6</b>	<b>PHASE 4 – INTERPRET RESULTS, MODIFY FORM 2 &gt; FORM 3</b>	<b>112</b>
	4.6.1 INSTRUCTIONS	112
	4.6.2 FLOW	112
	4.6.3 METHODOLOGY	112
	4.6.4 RESULTS	112
	4.6.5 INSTRUCTIONS	112
	4.6.6 PROBLEM IDENTIFICATION	112
	4.6.7 METHODOLOGY	112
<b>4.7</b>	<b>SUMMARY</b>	<b>113</b>

## CHAPTER 5 – CONDUCT OF THE STUDY

<b>5.1</b>	<b>INTRODUCTION</b>	<b>114</b>
	5.1.1 HYPOTHESIS	114
	5.1.2 POPULATION/METHODOLOGY	114
	5.1.3 INSTRUMENT	114
	5.1.4 REFINEMENT AND TEST	114
	5.1.5 REFINEMENT	114
	5.1.6 IMPLEMENTATION	114
<b>5.2</b>	<b>PHASE 5 – USE FORM 3</b>	<b>115</b>
<b>5.3</b>	<b>PHASE 6 – INTERPRET RESULTS, DECISION ON MODEL</b>	<b>117</b>
	5.3.1 SEARCH ENGINES	118
	5.3.2 STRATEGY	118
	5.3.3 SPEED	120
	5.3.4 CHOICE OF SEARCH ENGINE	120
	5.3.5 OPERATORS	121
	5.3.6 SPEED	122
<b>5.4</b>	<b>PHASE 7 – TEST FORM 3 WITH MODEL 1</b>	<b>123</b>

<b>5.5</b>	<b>PHASE 8 – INTERPRET RESULTS, MODIFY MODEL 1 &gt; MODEL 2</b>	<b>123</b>
	5.5.1 COUNTRY DEPENDENCE/SLEEPING SERVERS	123
	5.5.2 QUERY	124
<b>5.6</b>	<b>PHASE 9 – TEST FORM 3 WITH MODEL 2</b>	<b>124</b>
<b>5.7</b>	<b>PHASE 10 – INTERPRET RESULTS, MODIFY FORM 3 &gt; FORM 4, MODEL 2 &gt; MODEL 3</b>	<b>125</b>
	5.7.1 QUERY	125
	5.7.2 SEARCH ENGINE CHOICE	125
	5.7.2.1 Search engine choice	127
	5.7.2.2 Searching methodology	129
	5.7.2.3 Search engine tests	130
	5.7.2.4 Search engine test results	144
<b>5.8</b>	<b>MODEL 3 (FOIOTI)</b>	<b>144</b>
	5.8.1 DESIGN CRITERIA	144
	5.8.1.1 Basic approach	145
	5.8.1.2 Experimental Methodology	145
	5.8.1.3 Query formulation	146
	5.8.1.4 Training session	146
	5.8.1.5 Multiple indices	147
	5.8.2 DESCRIPTION OF OPERATION	147
	5.8.3 EXAMPLE	149
<b>5.9</b>	<b>PHASE 11 – TEST FORM 4 WITH MODEL 3</b>	<b>150</b>
<b>5.10</b>	<b>PHASE 12 – INTERPRET RESULTS, DETERMINE SUCCESS OF MODEL 3</b>	<b>151</b>
<b>5.11</b>	<b>PHASE 13 – IMPLEMENT MODEL 3 ON CONTINUOUS BASIS</b>	<b>154</b>
<b>5.12</b>	<b>CLASS 4 RESULTS</b>	<b>156</b>
	5.12.1 RESULTS	156
	5.12.2 ANALYSIS	157
	5.12.2.1 Wide specification	158
	5.12.2.2 Time usage	158
	5.12.2.3 Multiple indices	158
	5.12.2.4 Conclusion	159

## CHAPTER 6 – RESEARCH RESULTS AND ANALYSIS

<b>6.1</b>	<b>INTRODUCTION</b>	<b>160</b>
<b>6.2</b>	<b>AGE</b>	<b>161</b>
	6.2.1 RESULTS	161
	6.2.2 ANALYSIS	162
<b>6.3</b>	<b>GENDER</b>	<b>163</b>
	6.3.1 RESULTS	163
	6.3.2 ANALYSIS	164
<b>6.4</b>	<b>RACE</b>	<b>164</b>
	6.4.1 RESULTS	164
	6.4.2 ANALYSIS	167
<b>6.5</b>	<b>VALIDITY</b>	<b>168</b>
	6.5.1 RESULTS	168
	6.5.2 ANALYSIS	168
<b>6.6</b>	<b>FORMAL INSTRUCTION</b>	<b>170</b>
	6.6.1 RESULTS	170
	6.6.2 ANALYSIS	171
<b>6.7</b>	<b>INTERNET USAGE</b>	<b>172</b>
	6.7.1 RESULTS	172
	6.7.2 ANALYSIS	173
<b>6.8</b>	<b>SEARCH ENGINE USAGE</b>	<b>174</b>
	6.8.1 RESULTS	174
	6.8.2 ANALYSIS	175

<b>6.9</b>	<b>SEARCH ENGINES</b>	175
	6.9.1 RESULTS	175
	6.9.2 ANALYSIS	178
<b>6.10</b>	<b>RESULTS RECEIVED</b>	179
	6.10.1 RESULTS	179
	6.10.2 ANALYSIS	181
<b>6.11</b>	<b>NUMBER OF WEBSITES VISITED</b>	182
	6.11.1 RESULTS	182
	6.11.2 ANALYSIS	184
<b>6.12</b>	<b>NUMBER OF YES ANSWERS BEFORE MODEL</b>	184
	6.12.1 RESULTS	184
	6.12.2 ANALYSIS	185
<b>6.13</b>	<b>TIME SPENT SEARCHING</b>	188
	6.13.1 RESULTS	188
	6.13.2 ANALYSIS	189
<b>6.14</b>	<b>YES ANSWERS AFTER MODEL 1, 2 AND 3</b>	190
	6.14.1 RESULTS	190
	6.14.2 ANALYSIS	191
<b>6.15</b>	<b>NUMBER OF USERS CHOOSING WEBCRAWLER/EXCITE/HOTBOT</b>	192
	6.15.1 RESULTS	192
	6.15.2 ANALYSIS	193
<b>6.16</b>	<b>NUMBER OF PARTICIPANTS PER CLASS</b>	194
	6.16.1 RESULTS	194
	6.16.2 ANALYSIS	195

## CHAPTER 7 – CONCLUSION

<b>7.1</b>	<b>INTRODUCTION</b>	196
<b>7.2</b>	<b>SUMMARY</b>	196
	7.2.1 DEMOGRAPHICS	196
	7.2.2 INTERNET EXPOSURE	197
	7.2.3 SEARCHING DETAIL	197
	7.2.4 SIGNIFICANCE	198
	7.2.5 FURTHER RESEARCH	200
	7.2.5.1 Focussed FOIOTI	200
	7.2.5.2 Hyper help	200
	7.2.5.3 Specification enhancer	204
	7.2.6 SEARCHING SUCCESS	205
<b>7.3</b>	<b>CONCLUSION</b>	205

## APPENDICES

<b>APPENDIX A</b>	-	<b>Form 1</b>	209
<b>APPENDIX B</b>	-	<b>Form 2</b>	211
<b>APPENDIX C</b>	-	<b>Form 3</b>	214
<b>APPENDIX D</b>	-	<b>Form 4</b>	217
<b>APPENDIX E</b>	-	<b>Form 5</b>	220
<b>APPENDIX F</b>	-	<b>Model 1</b>	222
<b>APPENDIX G</b>	-	<b>Model 2</b>	224
<b>APPENDIX H</b>	-	<b>Screendump of Model 3 (FOIOTI)</b>	226
<b>APPENDIX I</b>	-	<b>Not used</b>	228

<b>APPENDIX J - Experiment Results</b>	<b>229</b>
<b>J.1 RESULTS AND ANALYSIS: PHASE 1 – DESIGN AND TEST INSTRUMENT: FORM 1</b>	<b>230</b>
J.1.1 AGE	230
J.1.2 GENDER	230
J.1.3 RACE	230
J.1.4 VALIDITY	231
J.1.5 SEARCH ENGINES	231
J.1.6 RESULTS RECEIVED	231
J.1.7 NUMBER OF WEBSITES VISITED	232
J.1.8 NUMBER OF YES ANSWERS BEFORE MODEL	232
J.1.9 TIME SPENT SEARCHING	232
<b>J.2 RESULTS AND ANALYSIS: PHASE 3 – TEST FORM 2 AS INSTRUMENT</b>	<b>232</b>
J.2.1 AGE	232
J.2.2 GENDER	233
J.2.3 RACE	233
J.2.4 VALIDITY	233
J.2.5 FORMAL INSTRUCTION	233
J.2.6 INTERNET USAGE	234
J.2.7 SEARCH ENGINES	234
J.2.8 RESULTS RECEIVED	234
J.2.9 NUMBER OF WEBSITES VISITED	235
J.2.10 NUMBER OF YES ANSWERS BEFORE MODEL	235
J.2.11 TIME SPENT SEARCHING	235
<b>J.3 RESULTS AND ANALYSIS: PHASE 5 – TEST FORM 3</b>	<b>235</b>
J.3.1 AGE	235
J.3.2 GENDER	236
J.3.3 RACE	236
J.3.4 VALIDITY	236
J.3.5 FORMAL INSTRUCTION	236
J.3.6 INTERNET USAGE	237
J.3.7 SEARCH ENGINE USAGE	237
J.3.8 SEARCH ENGINES	238
J.3.9 RESULTS RECEIVED	238
J.3.10 NUMBER OF WEBSITES VISITED	239
J.3.11 NUMBER OF YES ANSWERS BEFORE MODEL	239
J.3.12 TIME SPENT SEARCHING	239
<b>J.4 RESULTS AND ANALYSIS: PHASE 7 – TEST FORM 3 WITH MODEL 1</b>	<b>239</b>
J.4.1 AGE	239
J.4.2 GENDER	240
J.4.3 RACE	240
J.4.4 VALIDITY	240
J.4.5 FORMAL INSTRUCTION	240
J.4.6 INTERNET USAGE	241
J.4.7 SEARCH ENGINE USAGE	241
J.4.8 SEARCH ENGINES	241
J.4.9 RESULTS RECEIVED	242
J.4.10 NUMBER OF WEBSITES VISITED	242
J.4.11 NUMBER OF YES ANSWERS BEFORE MODEL	242
J.4.12 TIME SPENT SEARCHING	242
J.4.13 NUMBER OF YES ANSWERS AFTER MODEL 1	243

<b>J.5</b>	<b>RESULTS AND ANALYSIS: PHASE 9 – TEST FORM 3 WITH MODEL 2</b>	<b>243</b>
J.5.1	AGE	243
J.5.2	GENDER	243
J.5.3	RACE	244
J.5.4	VALIDITY	244
J.5.5	FORMAL INSTRUCTION	244
J.5.6	INTERNET USAGE	245
J.5.7	SEARCH ENGINE USAGE	245
J.5.8	SEARCH ENGINES	245
J.5.9	RESULTS RECEIVED	246
J.5.10	NUMBER OF WEBSITES VISITED	246
J.5.11	NUMBER OF YES ANSWERS BEFORE MODEL	246
J.5.12	TIME SPENT SEARCHING	246
J.5.13	NUMBER OF YES ANSWERS AFTER MODEL 2	247
<b>J.6</b>	<b>RESULTS AND ANALYSIS: PHASE 11 – TEST FORM 4 WITH MODEL 3</b>	<b>247</b>
J.6.1	AGE	247
J.6.2	GENDER	247
J.6.3	RACE	248
J.6.4	VALIDITY	248
J.6.5	FORMAL INSTRUCTION	248
J.6.6	INTERNET USAGE	249
J.6.7	SEARCH ENGINE USAGE	249
J.6.8	SEARCH ENGINES	250
J.6.9	RESULTS RECEIVED	250
J.6.10	NUMBER OF WEBSITES VISITED	251
J.6.11	NUMBER OF YES ANSWERS BEFORE MODEL 3	251
J.6.12	TIME SPENT SEARCHING	251
J.6.13	NUMBER OF USERS CHOOSING WEBCRAWLER	251
J.6.14	NUMBER OF USERS CHOOSING EXCITE	252
J.6.15	NUMBER OF USERS CHOOSING HOTBOT	252
J.6.16	NUMBER OF WEBSITES VISITED	252
J.6.17	AMOUNT OF TIME SPENT	252
J.6.18	NUMBER OF YES ANSWERS AFTER MODEL 3	253
J.6.19	NUMBER OF PARTICIPANTS PER CLASS	253
J.6.20	NUMBER OF PARTICIPANTS PER CLASS (ADJUSTED - 5)	253
J.6.21	NUMBER OF PARTICIPANTS PER CLASS (ADJUSTED - 2)	253
<b>J.7</b>	<b>RESULTS AND ANALYSIS: PHASE 13 – IMPLEMENT MODEL 3</b>	<b>254</b>
J.7.1	AGE	254
J.7.2	GENDER	254
J.7.3	RACE	254
J.7.4	VALIDITY	254
J.7.5	FORMAL INSTRUCTION	255
J.7.6	INTERNET USAGE	255
J.7.7	SEARCH ENGINE USAGE	255
J.7.8	SEARCH ENGINES	256
J.7.9	RESULTS RECEIVED	256
J.7.10	NUMBER OF WEBSITES VISITED	256
J.7.11	NUMBER OF YES ANSWERS BEFORE MODEL 3	257
J.7.12	TIME SPENT SEARCHING	257
J.7.13	NUMBER OF USERS CHOOSING WEBCRAWLER	257
J.7.14	NUMBER OF USERS CHOOSING EXCITE	257
J.7.15	NUMBER OF USERS CHOOSING HOTBOT	258
J.7.16	NUMBER OF WEBSITES VISITED	258
J.7.17	AMOUNT OF TIME SPENT	258
J.7.18	NUMBER OF YES ANSWERS AFTER MODEL 3	258
J.7.19	NUMBER OF PARTICIPANTS PER CLASS	259
J.7.20	NUMBER OF PARTICIPANTS PER CLASS (ADJUSTED - 5)	259
J.7.21	NUMBER OF PARTICIPANTS PER CLASS (ADJUSTED - 2)	259

<b>APPENDIX K</b>	-	<b>Examples</b>	260
<b>APPENDIX L</b>	-	<b>Source Code of Model 3 (FOIOTI)</b>	264
<b>APPENDIX M</b>	-	<b>Statistical References</b>	280
<b>BIBLIOGRAPHY</b>			289
<b>GLOSSARY</b>			319

## TABLES

<b>TABLE 1.1</b>	-	<b>Extract from URL file</b>	13
<b>TABLE 2.1</b>	-	<b>Growth of the number of Internet Hosts</b>	33
<b>TABLE 4.1</b>	-	<b>Interaction of Sample and Population Size</b>	99
<b>TABLE 4.2</b>	-	<b>Phase 1 Session 1 - 5</b>	108
<b>TABLE 4.3</b>	-	<b>Phase 3 Session 6</b>	111
<b>TABLE 5.1</b>	-	<b>Phase 5 Session 7 – 21</b>	116
<b>TABLE 5.2</b>	-	<b>Yes/No answers Phase 1, 3, 5</b>	117
<b>TABLE 5.3</b>	-	<b>Phase 6 Error Frequency</b>	120
<b>TABLE 5.4</b>	-	<b>Chosen Search Engines</b>	121
<b>TABLE 5.5</b>	-	<b>Phase 7 Session 22 – 24</b>	123
<b>TABLE 5.6</b>	-	<b>Phase 9 Session 25 – 27</b>	124
<b>TABLE 5.7</b>	-	<b>Search Engine Choice</b>	128
<b>TABLE 5.8</b>	-	<b>Highest Search Engine Scores</b>	129
<b>TABLE 5.9</b>	-	<b>Operators used by Searchers</b>	130
<b>TABLE 5.10</b>	-	<b>Results from Request 1</b>	133
<b>TABLE 5.11</b>	-	<b>Results from Request 2</b>	134
<b>TABLE 5.12</b>	-	<b>Results from Request 3</b>	134
<b>TABLE 5.13</b>	-	<b>Results from Request 4</b>	135
<b>TABLE 5.14</b>	-	<b>Results from Request 5</b>	136
<b>TABLE 5.15</b>	-	<b>Results from Request 6</b>	137
<b>TABLE 5.16</b>	-	<b>Results from Request 7</b>	137
<b>TABLE 5.17</b>	-	<b>Results from Request 8</b>	138
<b>TABLE 5.18</b>	-	<b>Results from Request 9</b>	138
<b>TABLE 5.19</b>	-	<b>Results from Request 10</b>	139
<b>TABLE 5.20</b>	-	<b>Search Engine Success</b>	139
<b>TABLE 5.21</b>	-	<b>Operator Category Success</b>	140
<b>TABLE 5.22</b>	-	<b>Results from Request 11</b>	140
<b>TABLE 5.23</b>	-	<b>Results from Request 12</b>	141
<b>TABLE 5.24</b>	-	<b>Results from Request 13</b>	141
<b>TABLE 5.25</b>	-	<b>Results from Request 14</b>	141
<b>TABLE 5.26</b>	-	<b>Results from Request 15</b>	142
<b>TABLE 5.27</b>	-	<b>Results from Request 16</b>	142
<b>TABLE 5.28</b>	-	<b>Results from Request 17</b>	142
<b>TABLE 5.29</b>	-	<b>Results from Request 18</b>	143
<b>TABLE 5.30</b>	-	<b>Results from Request 19</b>	143
<b>TABLE 5.31</b>	-	<b>Results from Request 20</b>	144
<b>TABLE 5.32</b>	-	<b>Search Engine Success</b>	144
<b>TABLE 5.33</b>	-	<b>Example of FOIOTI's Operation</b>	149
<b>TABLE 5.34</b>	-	<b>Phase 11 Session 28 – 38</b>	151
<b>TABLE 5.35</b>	-	<b>FOIOTI class Definition</b>	152
<b>TABLE 5.36</b>	-	<b>FOIOTI Class 5 Reasons</b>	152
<b>TABLE 5.37</b>	-	<b>Phase 12 – FOIOTI Class results (5x)</b>	153
<b>TABLE 5.38</b>	-	<b>Phase 12 – FOIOTI Class results (3x)</b>	153
<b>TABLE 5.39</b>	-	<b>Phase 13 Session 39 – 46</b>	154
<b>TABLE 5.40</b>	-	<b>FOIOTI Class results (5x)</b>	155
<b>TABLE 5.41</b>	-	<b>FOIOTI Class results (3x)</b>	155
<b>TABLE 5.42</b>	-	<b>FOIOTI Class results (3x)</b>	156

<b>TABLE 6.1</b>	-	<b>Phase Summary</b>	160
TABLE 6.2	-	Age Summary	161
TABLE 6.3	-	Gender Summary	163
TABLE 6.4	-	Race Summary	166
TABLE 6.5	-	Validity summary	168
TABLE 6.6	-	Formal Internet Instruction Summary	170
TABLE 6.7	-	Internet Usage Summary	172
TABLE 6.8	-	Search Engine Usage Summary	174
TABLE 6.9	-	Search Engine Summary	177
TABLE 6.10	-	Results Received Summary	180
TABLE 6.11	-	Number of Websites Visited Summary	183
TABLE 6.12	-	Number of Yes Answers before Model Summary	185
TABLE 6.13	-	Time Spent Searching Summary	188
TABLE 6.14	-	Yes Answers after Model 1, 2, 3 Summary	190
TABLE 6.15	-	Number of Users choosing Search Engines Summary	192
TABLE 6.16	-	Number of Participants per Class Summary	194
<b>TABLE 7.1</b>	-	<b>Hyper Help Example</b>	203

## GRAPHS

<b>GRAPH 2.1</b>	-	<b>Internet Host growth Period 1</b>	34
GRAPH 2.2	-	Internet Host growth Period 2	34
<b>GRAPH 6.1</b>	-	<b>Age Summary</b>	162
GRAPH 6.2	-	Gender Distribution	164
GRAPH 6.3	-	Race Distribution	166
GRAPH 6.4	-	Formal Internet Instruction Summary	171
GRAPH 6.5	-	Internet Usage Distribution	173
GRAPH 6.6	-	Search Engine Usage summary	175
GRAPH 6.7	-	Search Engine Summary	178
GRAPH 6.8	-	Results Received Summary: Average	180
GRAPH 6.9	-	Results Received Summary: Maximum	181
GRAPH 6.10	-	Number of Websites Visited Summary	183
GRAPH 6.11	-	Time spent Searching Summary	189
GRAPH 6.12	-	Comparison of Yes Answers	191
GRAPH 6.13	-	Number of Users choosing Search Engines Summary	193
GRAPH 6.14	-	Number of Users per Class – FOIOTI	194

# **CHAPTER 1**

## **BACKGROUND AND PROBLEM STATEMENT**

### **1.1 INTRODUCTION**

This research project focuses on Information Technology and Information Systems (IT/IS) learners, and an important problem they face in their learning process: resources. The author describes this problem, identifies a potential solution, explains how it was designed, tested and implemented. The aim of this project was to harness the power of the Internet as an information source by involving the learner in his/her own educational progress. As such it is concerned with "information literacy", which has been defined as:

"... the ability of learners to access, use and evaluate information from different sources, in order to enhance learning, solve problems and generate new knowledge" (Adamastor 2001).

This project focuses on the aspect of information literacy principally concerned with the finding of relevant information. The term "the author" will be used to refer to the author of this research project, while "this author" will refer to another author whose work is being discussed at that point.

In the early 1980s, computers in the form of personal computers (PCs) started making their mark in the business world and in education. The IT/IS learner initially worked on a dumb terminal where he/she had no control over any external components except to execute some programs and view the results on a monitor.

Currently, however, the learner operates in a volatile environment where he/she has extensive control over programs and information residing in other computers thousands of kilometres away, communication networks, multimedia, the Internet and even other users via a PC.

## **1.2            THE LEARNING SPACE**

### **1.2.1            SCHOOLING**

The primary and secondary education system in South Africa involves 12 years of schooling, from grade 1 to 12. After grade 12 (also called matric), school leavers are ready to either enter the job market, or enrol in one of many courses offered by the higher education institutions. Unfortunately, one-way communication and rote learning, partially due to a lack of resources, sometimes characterise learning during these 12 years of schooling. Potenza describes this process as "The silent hell of having to listen to teachers droning on endlessly..." (Potenza and Jansen 1998:54). Some learners exit the schooling system where lack of resources and uninterested educators discourage any experimentation with resource-based teaching.

### **1.2.2            INSTITUTIONS OF HIGHER EDUCATION**

South Africa offers a school leaver having just completed matric and interested in higher education, many types of institutions to choose from. Universities and technikons form the backbone of this choice: 75% of all higher education students are registered at one of these institutions (NCHE 2000). In the Western Cape (one of the nine provinces of South Africa), five of these Institutions exist:

- The University of Cape Town
- The University of Stellenbosch
- The University of the Western Cape
- Peninsula Technikon
- The Cape Technikon.

Universities mostly specialise in academic and theoretical courses with practical components, typically offering the following qualifications:

Bachelor's degree (for example B.Comm. and B.Sc.)	-	3 years
Honours degree	-	1 year
Master's degree		
Doctorate		

Technikons focus on practice-oriented training, often including co-operative oriented education, and typically offer the following qualifications:

National Diploma (for example ND in IT and ND in Financial Information Systems)	-	3 years
B. Tech degree	-	1-2 years
M. Tech degree		
D. Tech degree		

The National Diploma in Information Technology (NDIT) consists of subjects such as: Information Systems I, II and III, Development Software I, II and III, Systems Software II and III, Technical Programming I, Financial Accounting I and IT Skills I. Each of these subjects could require four to six 45-minute lectures per week, for an average of 32 weeks of lecturing per academic year. Half of the lectures could be typical, traditional theoretical lectures and the remainder practical hands-on sessions in a computer laboratory. The contents of one of these subjects (Systems Software II for example) are: computer architecture, data communications, operating systems (all theory) and low-level programming, networking and Internet work (all practical).

### 1.2.3 THE LEARNERS

A diplomate from the Cape Technikon NDIT course could be employed in the IT industry as a Programmer or a Systems Analyst. Languages used vary from the more traditional COBOL, Natural and Visual Basic to more modern types such as HTML, Java, C++ and Object Oriented platforms.

Approximately one out of every five applicants for the IT course is accepted, based on the maximum intake possible due to laboratory size and personnel constraints. The profile of a typical first-year student in the IT course is as follows:

Age:	19 years.
Race:	White, Coloured or African.
Gender:	Mostly male.
Highest qualification:	Matric (grade 12).
Previous experience:	Some have done computer science at school.

The author has been involved in computing for two decades, of which the last 16 years have been spent in educating learners in a variety of subjects at the Faculty of Business Informatics, Cape Technikon, mostly towards the NDIT mentioned above.

#### **1.2.4 EDUCATION IN SOUTH AFRICA: CURRENT SITUATION**

The National Qualifications Framework (NQF) was established in South Africa to redress the educational inequalities of the past. The objectives of the NQF are to build a seamless national learning system, and to restore equity to the South African higher education system (NAPTOSA 2000:2). Eight levels were defined, encompassing education from pre-school to the Senior school phase (level 1), school and College training (levels 2 to 4) to levels 5 to 8 (diplomas, first degrees, higher degrees and doctorates). It was proposed that the NQF become a reality via two legs: Recognition of Prior Learning (RPL) and Outcomes Based Education (OBE).

RPL involves recognition of relevant work experience for entrance to higher education, as opposed to strict adherence to rules and regulations regarding prior qualifications or levels of education achieved (Gawe 2001).

Killen describes OBE as: "... an approach that requires teachers and students to focus their attention and efforts on the desired end results of education" (Killen 1997:25). OBE has been shrouded in controversy since it was announced, as described by both local and international authors (Manno 1994:1, Potenza and Jansen 1998:54, Spady 1994:141). However, OBE is currently being implemented in South African schools, originally under the title Curriculum 2005, later renamed Curriculum 21.

One crucial element of the OBE system is the fact that educators should "... select appropriate content..." (Killen 1997:32). Potenza and Jansen claim that educators will no longer have to use a specific textbook, but will have the freedom to choose from a wide range of resources (Potenza and Jansen 1998:56). OBE is clearly a resource-based learning system. It is this element of finding the best resources, not only by school but also higher education learners, which is addressed in this study. It will be shown in Chapter 2 (see Section 2.2.3) that the casual use of the many available search engines on the Internet seldom leads to success in information retrieval. If this problem cannot be solved, the freedom of choice of resources might become more of a problem than being part of a solution.

### **1.3 MOTIVATION**

While working with IT/IS learners, a number of realities became evident. The author soon realised that learning does not necessarily take place if some ingredients (a classroom, books, students and a lecturer) are combined and stirred. The author initially assumed that a 19-year old learner enrolling for a three-year IT course, and paying a large sum of money to be allowed to do so, would automatically be motivated to participate in classes and certain to achieve overall success. This assumption was confirmed as an expectation by Killen (1997:27). However, the author was not correct in this assumption, which embodies the first major problem noticed by the author. A question that naturally arose was: **how could learners be motivated to put more into their studies?**

#### 1.4 LEARNER PARTICIPATION

A second problem identified was the lack of learner involvement in classes. This was not only limited to theory classes, but also appeared in practical laboratory sessions. In 1995 the author launched an informal, undocumented experiment to determine whether it is possible to convince learners to participate in their own progress. They were second-year students enrolled for the three-year NDIT. The subject was Systems Software II, which involved the topics as described in Section 1.2 above. The learners were given some advice on how to and how not to set multiple choice questions.

Three weeks before an official test on the subject, they were asked to compile and submit their own multiple choice questions, covering all work to be evaluated in that test.

The learners were promised that the lecturer would evaluate these questions, and include the ones he considered to be "good" questions in the test. This implied that if a learner participated in this experiment, he/she would stand a chance of answering some questions in a test, which he/she had set himself/herself.

The response was overwhelming. Although only approximately 50% of the submitted questions were eventually included in the test, it proved a point: learners will participate in a class when their own evaluation is at stake. The author thus realized that this result indicated that the potential exists to solve the participation problem by using learner involvement. The question arose: **which method(s) could be used (on a regular basis) to achieve this co-operation between educator and learner, to the learner's advantage?** This question is central to the theme of this research project.

#### 1.5 RESOURCES

In the same vein, it was learnt that the presence of a textbook, handouts, a library or a computer did not guarantee that learning would take place from these resources. This became clear after the following sequence of events occurred a number of times:

- the author covered certain work in theory
- he identified resources to be used by the learner to reinforce the work covered, sometimes including handouts provided by the author
- evaluation was done on the same work
- from the results produced by the learners, it was clear that the resources had not been consulted.

The author identified a perception among many learners that as long as the educator produced notes, lectures, examples and references, and as long as the learner ensured ownership of these resources, during or after the class, education was taking place and subsequent gainful employment was imminent.

The third major problem identified was thus that the presence of resources was not enough to guarantee that learners would use them effectively. As a result, the author asked himself a third question: **why did learners not make use of the resources provided for their own benefit?**

### **1.5.1 USE OF RESOURCES**

The author found that a fair amount of time in a typical day was spent on learner resources. These included:

- using prescribed textbooks to prepare new work
- evaluating new textbooks to be able to make informed choices every year about new prescribed textbooks for the next academic year for each subject
- typing/photocopying and distributing handouts
- designing, then writing or typing and printing transparencies
- managing handouts and transparencies on the intranet to allow learners access to them
- management of paper-based resources in the library short-loan section.

As a result, it was considered necessary to discuss these resources in more detail.

### 1.5.1.1 TEXTBOOKS

A lecturer normally has a choice of which resources to use in a subject, and their format. Traditionally a textbook is prescribed, which the learner has to purchase and use for each subject. However, the publishing chain, which includes the process of peer review, commissioning, marketing, printing and distribution cannot produce new texts fast enough for the changes in the Information Technology field. The author has found the delays associated with traditional printed matter to be unacceptable, owing to the rapid growth in Information Technology.

Furthermore, some educators have been accused of prescribing expensive textbooks, which are then seldom used because the contents are outdated, or because the content is not adequate for the wide scope of work being covered in one subject. This is especially true in subjects where the contents are offered as a set of modules, for example: Information Systems III is divided into Systems Analysis and Design, Databases and Project Management. Each module is typically offered by a different lecturer, and could require that a complete textbook costing upwards of R300 be prescribed. A similar problem exists with computer program manuals: the rapid release of new versions of computer programs has made the purchase of large numbers of manuals by educational institutions an expensive exercise.

It appeared that the problem with these resources could be resolved if a resource was found which could be updated easily, or even better, was updated automatically. If it could also obviate the need for purchasing expensive books, it would be a major advantage.

### 1.5.1.2 HANDOUTS AND TRANSPARANCIES

As a result of outdated textbooks, or their lack of the exact coverage of material desired by the lecturer, handouts and/or transparencies are often used to supplement textbooks. This evinced a new set of problems:

- the copyright issue - is it permissible to make copies of sections of a book? How many pages may be copied before it becomes illegal? A body exists (The Adamastor Trust Publishing Liaison Office) to serve the five Higher Education Institutions in the Western Cape mentioned in Section 1.2.2, in this regard. It warns that fines of up to R10 000 and/or imprisonment of up to five years are possible for the contravention of the Copyright Act (Van der Merwe 2000:1). Clear instructions are given on what may be legally reproduced for teaching purposes, but these instructions are difficult to apply in all situations.

For example, one instruction states that:

"... 'not more than one short poem, article, story or essay or two excerpts copied from the same author or more than three short poems, articles, stories or essays from the same collective work or periodical volume' and 'no more than 9 instances of such multiple copying for one course of instruction to a particular class during any one term' may be made without copyright permission" (Van der Merwe 2000:1).

Furthermore, if copyright permission is required, it could take up to two months to receive a response for overseas requests.

Some educators work with more than 300 learners per week, spread over three or four different courses of instruction. Keeping record of the number of articles copied per term per course and waiting up to two months for permission to copy one single article is an impossible task for an educator in the IT field. The actual content of a course may be changed at short notice, owing to new developments, and this kind of time frame is simply too restrictive.

- management – the preparation, duplication, distribution, storage and evaluation of handouts and transparencies can become an onerous task.
- time – it takes time to design, create and produce transparencies or even just duplicate handouts.

The author has found that even though the work covered by a handout or transparency is familiar, a certain amount of learning takes place while the medium is being prepared. However, the most serious drawback of these two types of resources is that this learning process is not automatically transferred when the resource is handed over to the learner. This is obvious when evaluation is later done on concepts covered by these materials. Often the author had an expectation of learning having taken place as a result of work having been covered using certain handouts and/or transparencies. When marking the tests or projects on this work, it was evident that this learning process had not taken place.

#### 1.5.1.3 THE LIBRARY

A library is, and will probably remain, an extremely important resource in any academic institution. The author found that learners were not keen to make use of resources specifically placed in the library for their use. This phenomenon was noted as being common to engineers as early as 1961. Jackson quoted Landis urging "... the engineering profession to increase its efforts to obtain effective use of published information" (Jackson 1971:431). Over the past years, various materials have been provided in the library short-loan section, where any library user could reserve it for periods for up to one day. The author has compared the number of stamps on the cover pages of these materials with the number of students in all the classes requiring them, and has found that only a small percentage of learners actually use this facility. This was confirmed when a test or project on these materials was marked – lack of insight was often evident. A longitudinal study done on engineering students' use of the library at their university indicated a similar trend of low usage patterns (Holland and Powell 1995:14).

#### 1.5.2 THE INTERNET

During the last few years Internet access in the Cape Technikon computer laboratories was established as a standard feature. It was initially viewed by most lecturers as a "must-have", but at the time of writing only two educators in the Faculty of Business Informatics, including the author, have integrated the Internet as an official resource in subjects being taught on a daily basis.

However, another problem surfaced at this point. As the Internet became more accessible in the computer labs, aimless surfing, chatting and e-mail became popular pastimes among learners. The author found that many learners did not do the practical work expected of them in the labs, not even during "supervised practicals".

Instead, they spent large amounts of time on the Internet reading e-mail and visiting sites which appeared to be unrelated to the work being covered at the time. The attraction of colourful websites with animation and sound, screens full of information and the ubiquitous pornography, took the place of programming and other required tasks.

The author could not help but classify the average IT/IS student as one of the "Mail-Web-Internet Generation" defined by Tsichritzis as follows: "This young generation basks in its reputation as the people who surf the nets the widest and the longest. They lack technical skills but are adept at playing with the interfaces. ... They are excellent in a wide range of knowledge but lack depth anywhere" (Tsichritzis 1997:53).

However, the author believed firmly that the Internet had tremendous potential as a learning resource in education, and launched another unofficial experiment during 1998 to investigate this potential. The objective was to direct the time and energy learners spent on aimless surfing into providing reference material for the subject he/she was studying. Although the detail was different to that of the first experiment (see Section 1.4), the overall aim was the same: to find a way to involve the learner in his/her own learning process, to his/her own advantage.

The author challenged his learners with a scheme which required them to "officially" use the Internet to find information. If successful, an implicit reward in the form of recognition of their skills was offered.

Initially the learners were given basic background on the Internet, and on searching techniques and addresses of websites on the Internet (these addresses are called Universal Resource Locators, or URL's). Two hours in total were spent on this background training.

After having covered certain topics of a subject using traditional resources (textbooks, handouts, transparencies, etc), a two-hour session was arranged in a computer laboratory. A list of topics was given (based on the work covered during the preceding few weeks), and learners had to research these topics using any search engine and any technique to find relevant information. The author circulated through the class, and was called to a workstation every time a learner found a website, which he/she considered to contain relevant information. The author then evaluated the site. If it was considered to be unacceptable, a short explanation of the reason was given, and the learner would proceed with another search. If the site did contain relevant information, the lecturer copied the URL onto a diskette, and asked the successful learner to add his/her name to the URL. The lecturer then added a short description of how the contents of the website were to be interpreted. The learner then carried on with the next topic, enjoying the reward of his/her website having being recognized as "good enough" by the lecturer, and knowing that his/her name would be associated with the successful website. This implied that the learner had mastered a certain level of searching skills and general Internet awareness.

After approximately 1.5 hours, the author copied the file containing the URLs and comments onto the Intranet, where all learners could access it. They were instructed to copy this file, and visit each site listed before the end of the lab session. Although this sounds impossible to do in the time allowed, each URL simply had to be copied and not typed into the browser's address field (using the Windows clipboard). Furthermore, only the first learner to access a certain site might have had to wait some time for the contents to download. All laboratories access the Internet via a separate server, whose main purpose is to store all the most recent websites accessed by users.

Subsequent accesses within a certain period of time would involve a transfer of this information from the memory of this server (called cache memory) to the user's workstation, without actually accessing the Internet. This process plays a major part in the decrease of waiting time by Internet users.

The information found at the relevant website then had to be copied to their own diskette or network directory, thereby ensuring that the same information was found by all concerned.

It was made clear at the beginning that learners had to be present in the class. They could not expect the information identified as relevant to be available days, weeks and certainly not months after the date of identification. This is a result of the transient nature of information on the Internet, and the fact that webpages could disappear or have their names changed without warning.

An extract from a typical file, containing some of the URLs found by learners, is given below.

<p><b>1. <u>Boolean Algebra</u></b>  <b>Jaco Roos.</b>  <a href="http://fiddle.ee.vt.edu/courses/ee2504/combinational/logic_01.html">http://fiddle.ee.vt.edu/courses/ee2504/combinational/logic_01.html</a>  Simplification - Using Algebraic manipulation, gate implementation, Karnaugh maps. A few detailed worked out examples - read together with the problems on p11 and p12 of my BA notes.</p> <p><b>2. <u>Karnaugh Maps</u></b>  <b>Michael Lorenc.</b>  <a href="http://www.shef.ac.uk/~phys/teaching/phy107/kmap.html">http://www.shef.ac.uk/~phys/teaching/phy107/kmap.html</a>  Steps, labeling and looping of Karnaugh Maps - large, clear pictures, easy to follow.</p>
---

**TABLE 1.1 – Extract from URL file**

In general this experiment yielded positive results in terms of student learning. It was found that learners enjoyed applying their (otherwise ignored by the lecturer) Internet skills to do something considered to be useful by the author. The information found by the learners eventually formed part of the study material and subsequent tests, and the final examination included references to it. The result of this experiment created the impression that learners would readily use the Internet for academic purposes if a reward was given which enhanced their standing in a group. It convinced the author that the potential of the Internet as information resource for learners existed, and that a method should be found to develop this potential in a structured way.

### **1.6            SUMMARY**

While the author did not generalize by claiming that the observations above were true for all learners, the following points were noted:

- Current IT/IS learners were exposed to new challenges and opportunities compared to learners of 20 years ago.
- New methods were needed to motivate learners to become involved in their own learning process.
- A perception existed among learners that the onus to learn is on the educator, not the learner (confirmed by Killen: Killen 1997:29).
- A natural tendency against participation in a class situation appeared to exist among learners.
- Learners overcame this tendency if their involvement was related to their assessment, or when it enhanced their standing in a group.
- The use of traditional resources such as textbooks, handouts, transparencies and the library had many drawbacks.
- The availability of Internet access in a computer laboratory distracted learners.
- Internet access could be applied to the learner's advantage.

## 1.7 CONCLUSION

In conclusion, some research questions from the text above were summarized:

### LEARNER PARTICIPATION

- Is it possible to devise an educational tool which would increase learner involvement?
- Which method(s) can be used to achieve co-operation between educator and learner with regard to participation, to the learner's advantage?
- How can learners be motivated to put more into their studies?

### RESOURCES

- Why do learners not make use of the resources provided for their own benefit?
- Can the educational tool mentioned above be used to overcome some of the disadvantages of traditional resources?

**If these questions could be answered, collectively or individually, the author believed that the potential of the Internet as information resource for learners could be mined and applied to the benefit of learners in general.**

# **CHAPTER 2**

## **LITERATURE REVIEW**

### **2.1 INTRODUCTION**

Some fundamental issues on which this research project rests, had to be identified to guide the literature survey. Based on the summary and conclusion of the previous chapter, the following were suggested:

- educational methods associated with resource-based learning
- learner involvement in learning
- retrieval of information from resources
- the need for a retrieval tool to be used by those learning by resource-based methods, and
- the role of the Internet and search engines.

Having identified these issues, the author concentrated his reading on the following five major topics:

- a brief review of the history and status quo of Information Science and Information Retrieval, the Internet and the use of IT in education;
- the role of the World Wide Web as educational source in the library;
- studies and other experiments on Information Retrieval and the behaviour of searchers;
- search engines; and
- searching methodologies.

These five topics are summarised in this chapter, and analysed in the next chapter.

## **2.2 HISTORY AND STATUS QUO**

### **2.2.1 INFORMATION SCIENCE AND INFORMATION RETRIEVAL**

#### **2.2.1.1 INTRODUCTION**

Ingwersen defines Information Retrieval as being "... concerned with the processes involved in the representation, storage, searching and finding of information which is relevant to a requirement for information desired by a human user" (Ingwersen 1992: 49). A perspective on the development of Information Retrieval follows, since it is of value to this study. Specifically three aspects of Information Retrieval (of particular significance here) are outlined: these are the use of technology, content representation and the human perspective. The term IR will be used from here onwards to refer to Information Retrieval and WWW for World Wide Web. A summary of the development of Information Retrieval and Information Science, as reported by noted authors follows.

#### **2.2.1.2 HISTORY OF IR**

Ingwersen identifies three stages in the development of information science: its birth during the pre-Second World War era, its attempts to find an identity and alliances in the 1960s and 1970s and its establishment as a discipline during 1977-1980 (Ingwersen 1992:1).

#### **BIRTH PERIOD**

During the "birth" period, an article was published at the end of the Second World War, in which Bush suggested the possibility of building an electronic store of knowledge (Bush 1945). A possible technical solution was noted and a suggestion that the need and solution could interact in a way similar to which current users with information needs interact with the WWW. Information Science was vaguely defined and an Artificial Intelligence solution was proposed.

After this initial paving of the way for “modern” Information Science, IR was defined by Calvin Mooers in terms that quickly became accepted:

“In 1951, researcher and businessman Calvin Mooers coined the term *information retrieval*, to describe the process through which a prospective user of information can convert a request for information into a useful collection of references. Information retrieval embraces the intellectual aspects of the description of information and its specification for search, and also whatever systems, techniques, or machines that are employed to carry out the operation” (Gupta and Jain 1997:71, Spink and Losee 1996:43).

Mooers was one of the pioneers of Information Science, with his attempt to commercialize IR via his Zatocoding system (ASIS 2001, TRAC 2001). He also phrased what has come to be known as Mooers’s Law: “People will resist information unless the price of not knowing it greatly exceeds the price of learning it” (Garfield 2001).

The first of three strands of study which became evident during the 1950s related to the way in which people judge the relevance of documents. Even Mooers’s definition hints at this fundamental issue: “... a useful collection of references ...” (Gupta and Jain 1997:71, Spink and Losee 1996:43). Some questions arose: when is a collection of documents useful and relevant? How could relevance be recognized and measured?

Mizzaro did an exhaustive study on the history of relevance by discussing 157 papers on this topic, covering 40 years of IR research (Mizzaro 1997). Some of the authors whose studies were discussed include Barry, Burgin, Ellis, Harter, Hersh, Janes, Kantor, Mizzaro, Park, Rees, Saracevic, Schamber, Su, Sutton and Vickery. The following (sometimes contrasting) facts were stated about relevance:

- relevance is better understood than before
- an important challenge is to define relevance properly
- there is still a lot of work to be done to grasp its meaning fully

- relevance is a necessary part of understanding human behaviour
- as a study field it is "... frustrating, provocative, rich and -undeniably-relevant..." (Schamber 1994:36).

Relevance is discussed in detail in a subsequent section (see Section 2.5.6).

The second strand of the present study relates to the way in which search vocabulary maps information need to the terms used in a database. The third strand focuses on the role that computers play in the IR process. It appeared that IR could bring order into the fast-growing world of information, and Baeza-Yates and Ribeiro-Neto identified a need for "... efficient tools to manage, retrieve and filter information from this database" (Baeza-Yates and Ribeiro-Neto 1999:367). However, some studies have shown that users were not all eager to become involved in the extraction of knowledge from an electronic source. Jackson found that engineers are reluctant to use information sources. The United Engineering Information Service failed to elicit financial support from the engineering profession to establish its services (Jackson 1971:430).

#### IDENTITY PERIOD

Valuable contributions were made to the IR field by authors such as Burgin, Fidel, Mizzaro, Saracevic, Spink, Su and others. Saracevic, for example, pointed out that the important notion of relevance sets Information Science apart from other related fields, such as librarianship and documentation (Saracevic 1975:323-324). In later works, relevance as discussed here by Saracevic, was recognised as a parameter of IR performance and was renamed "precision". However, Saracevic's discussion focuses more on the perception of the user about the retrieval parameters of documents than on actual measurements of these parameters.

#### DISCIPLINE ESTABLISHMENT PERIOD

Lancaster predicted that information science might one day become a profession in its own right (Lancaster 1978:vii). During the next few years, a number of prominent authors started defining a number of concepts, which have become standard terminology in IR in general.

Marchionini, for example, defined the following relations between some key concepts: IR is a partial subset of information seeking, which is a partial subset of the learning process (Marchionini 1995:9). Just as IR forms part of Information Science, relevance is a crucial component of IR.

### MODERN PERIOD

The advent of the Internet and the WWW in the early 1990s created an information explosion unlike anything experienced by the industry before. IR appeared to be the tool to manage this situation to the advantage of the user (Saracevic 1999:1058). However, the rapid increase in the size of the Internet and therefore the amount of information available has put pressure on IR to become the solution to the growing problem of locating relevant information. An attempt to integrate several areas of Artificial Intelligence into an information gathering agent, called BIG (resource-Bounded Information Gathering) is described by Lesser and others. This agent manages to gather information to support a decision process, and reasons about the trade-offs of different resource gathering approaches. It is further capable of extracting information from structured and unstructured documents, using this information to refine its processing and search abilities (Lesser and others 2000:197).

A study done on the value of Internet research stated that the traditional rules of research and scholarship have changed irrevocably with the Internet now being used as a reference source. This author queried the standard of the information now so freely available (Nigohosian 1997).

Three aspects of the growth of IR, which has led to the current status of the field, have emerged from background reading. The first one is technological in nature, and is concerned with the means whereby the match between search terms and database entries are made. The second area of development is associated with the problem of representing the contents of a document in a retrieval system.

Finally the problem of psycholinguistics became evident: what happens in the mind of the searcher when an information need is transformed into a search strategy, and how does a searcher recognise the answer to his/her information need? These three aspects will be considered separately below.

### 2.2.1.3 TECHNOLOGICAL BACKGROUND

Firstly, work in the technological area started almost as soon as library collections were formed. This occurred approximately at the beginning of the third millennium BC. The Sumerians are credited as being the first people to store and classify written materials into "library collections", with the purpose of allowing various social groups to function better (Cover 2001). Everyday activities and literature were recorded on clay tablets, which were stored in special areas, with only a label bearing the opening words of the document as sole method of indexing (Greek Libraries 2001). The physical creation of these clay labels could be viewed as one of the first implementations of technology towards establishing indexing. In the absence of any advanced technological tools to make IR possible, these "libraries" were little more than marked collections of "documents".

The activities recorded by the Sumerians include administrative, political, judicial and social decisions, religious convictions and philosophical concerns. Even hymns, poems and myths were recorded in this way. Most importantly, detail about when the earliest schools began to function was recorded – including "tablet-repositories", which could be classified as the first libraries.

During the Middle Ages, indexing and simple classification of manuscripts were done. The indexers involved in this task were surrounded by an aura of mysticism, resulting from the coding schemes and alphabetical keys (Ingwersen 1992:61).

Work on cataloguing started after this period, using written, guardbook and card catalogues, followed by the period of mechanization.

This phase saw various card systems being developed, mostly in the form of punched cards: optical coincidence, the Zatocoding system of Mooers, edge-notched coding and 80-column punched cards.

Lancaster broadly defined the development of methods for the physical implementation of IR systems as spanning a number of decade-long periods:

- pre-1940s pre-co-ordinate systems, manual, printed book/card form
- 1940s post-co-ordinate systems
- 1950s punched cards, microfilm
- 1960s computer-based magnetic tape systems
- 1970s online computer based systems
- 1990s completely paperless systems (Lancaster 1978:15).

The high-powered computer era we currently find ourselves in provided the much-needed technology to empower this development, where document matching is made through inverted indices, string and positional searches. This increasing power of the technology has removed the economic constraint on the searching mechanism: any characteristic of the document could now be matched to a search query. In fact, there was now no technical constraint to prevent an index from including every single term of a given textual document in the index. This has been done some time ago, in the form of typical Bible concordances such as *Strong's Exhaustive Concordance*, first published in 1890. Since all the matching features so far have been based on text only, there remain some areas where matching has not yet been perfected: special notations such as music files, photographs and other graphical information.

#### 2.2.1.4 CONTENT REPRESENTATION BACKGROUND

Secondly, the content representation issue is addressed. The most ideal representation is simply to include the whole document as the index (as mentioned above), but the initial absence and later the limitations of technological tools (such as storage space) made this ideal impossible to achieve.

In the absence of computer-based full-text retrieval systems, two main lines of thought were initially followed to provide a summary of the document: assigned and derived indexing. Assigned indexing depended on an indexer's perception of the document contents, the matching of that perception to a term list or classification and the selection of appropriate descriptors to be assigned to the document. Derived indexing depended on the indexer (later the computer program) to extract terms as used in the document – keyword in context is the most notable early example.

During the late 1950s and 1960s, landmark work was done by noted authors in this area of document content presentation. The controversial Uniterm system sparked interest in the UK and the USA, which led to the Cranfield tests by Cleverdon and Keen (Cleverdon, Mills and Keen 1966, Robertson 2001, Tonta 2001).

Documents were indexed via a single term in the Uniterm system (hence the name), having been extracted from the document title or abstract. After some structured tests, Uniterm results were compared to those using more traditional indexing methods. The test apparently broke down owing to the disagreement over relevance judgement, and the results were inconclusive. One group of testers claimed that the Uniterm system worked well, while the other claimed the exact opposite (Ellis 1996:1,2).

The actual Cranfield series of tests was done at the College of Aeronautics, Cranfield, UK. It involved another comparison of performance, this time between the Uniterm system and a modified version of the Universal Decimal Classification (UDC) system. A subset of 200 documents on aeronautics was extracted and used as the master collection of documents. A further extract of 40 documents from this set was made, and these documents were used to generate a set of 40 artificial requests. The assumption was that if query number one was submitted, document number one should be returned as being the most relevant document from the collection of 200 documents.

Although certain inherent limitations of the study were evident, it did prove the effectiveness of the Uniterm system above the UDC classification (Ellis 1996:2).

This test was followed by the so-called Cranfield I test, which involved a comparison between four different indexing systems. Again a collection of master documents was used (this time 18 000), and it was indexed using each one of the four indexing systems. Using the same source document principle as with the previous test, 1200 questions were drawn up based on these documents. Searches were done, and the success or failure for each one noted. The performance of the four indexing systems was approximately equivalent, with a success level of between 73.8% and 82.0%. Again it was proved that the Uniterm system performed equally well compared to more conventional systems (Ellis 1996:3,4). Many authors levelled extensive criticism against the Cranfield I tests, mostly resulting from the concept of using source documents from which to derive the queries (Ellis 1996:5).

Another series of tests, aptly called Cranfield II, was undertaken next. It was based on different indexing devices/languages, rather than a test of indexing systems. A total of 211 search requests were elicited from the original authors of each document (out of a sample of 1400 documents). Thus the measure of effectiveness of this test compared to Cranfield I was explicitly relevance-based. This test proved that single-term index languages delivered the best performance, and that precision and recall are inversely related (Ellis 1996:6,7).

A number of other developments followed: the SMART retrieval system, Oddy's THOMAS retrieval program, Rich's GRUNDY system and a variety of expert intermediary systems like CONIT, CANSEARCH and PLEXUS (Ellis 1996:23,45,56,66).

Shortly after the Cranfield tests, Lancaster defined IR as follows: "Information retrieval is the process of searching some collection of documents ... in order to identify those documents which deal with a particular subject" (Lancaster 1978:11).

This author also claimed that the early history of computer-based IR had never been documented properly, and started work on MEDLARS - Medical Literature Analysis And Retrieval System. It was the first large, freely available system which allowed access without security or other restrictions (Lancaster 1978:35,36, Tonta 2001).

Two other studies worth mentioning are Blair and Maron's full-text retrieval system study, and the Dewey Decimal Classification online project of Markey and Demeyer. Blair and Maron used precision and recall as performance measures in their study on a 40 000 document database, based on 51 queries from two lawyers. They proved that recall failures occurred much more frequently than expected, and that the mean precision and mean recall ratio were 79% and 20% respectively. Markey and Demeyer viewed the Dewey Decimal Classification system as an online searcher's tool, and concluded that there was no relationship between the precision of online searches and the search satisfaction of the participants (Tonta 2001).

#### 2.2.1.5 HUMAN PERSPECTIVE BACKGROUND

In the third issue the focus moves from technology and methodology to the human side. How does the human searcher approach the task of finding information in a collection of information? How does he/she convert the perceived information need into a search strategy? What effect does experience have on the user's perception of information usefulness?

Vroom mentioned the concept of Expectancy, involving choices made by an individual. When the outcomes of a choice to be made are uncertain, his/her behaviour is affected not only by his/her own choice among these outcomes, but also by the degree to which he/she believes these outcomes to be probable (Vroom 1964:17).

Tonta reported on IR systems, choosing to focus on failure rather than success of document retrieval. Referring to the complexity of success or failure measurement in IR, Tonta admits that it is not even clear what a search failure is, and refers to user satisfaction as an alternative measurement.

The possibility of using precision as a measure of search failure is mentioned, where it is taken to be the failure to retrieve relevant documents only. Finally, three tools are defined which could be used to measure effectiveness of document retrieval:

- precision failures (retrieving non-relevant documents)
- recall failures (missing relevant documents), and
- fallout failures (retrieving too many irrelevant records) (Tonta 2001).

Ingwersen quotes Taylor, who postulated that a particular psychological state of mind of the user could result in an information need being expressed. Taylor identifies four levels of question formation, all defined as different types of needs:

- the visceral need (real but yet to be expressed need)
- the conscious need (brain-based, conscious description of the need)
- the formalised need (formal statement of the need)
- the compromised need (the information request as presented to the system) (Ingwersen 1992:112,113).

#### 2.2.1.6 STATUS QUO

The status quo of both Information Science and IR is best summarised by Saracevic when pondering on fifty years of development in this field (Saracevic 1999:1051). Three of its general characteristics were identified:

- Information Science spans various disciplines,
- Information Science is inherently connected to information technology,  
and
- Information Science plays an active part in the evolution of the information society.

Ingwersen identified a number of other disciplines which had an influence on Information Science: mathematics and communication (with an overlap in information theory), epistemology, sociology and linguistics (with an overlap in socio-linguistics), psychology, artificial intelligence and computer science (Ingwersen 1992:8).

Three fundamental “powerful ideas” on which information science is based are also listed: IR, relevance and interaction (Saracevic 1999:1052). This author thus clearly classifies IR as being part of Information Science.

## **2.2.2 THE INTERNET AND HYPERLINKS**

### **2.2.2.1 HISTORY OF THE INTERNET**

During the 1950s and 1960s, IR normally involved relatively static environments, including an organised collection of information, as found in libraries and Online Public Access Catalogues (OPACs) (Spink and Saracevic 1997:741). However, the advent of the Internet offered the user a larger mass of information, stored on computers spread across the globe, without a central catalogue or standardised system of indexing (Devlin and Burke 1997:102, Sha 1995:467, Stubenrauch, Vickery and Ruppert 1999). This contrast between the traditional and the new, the static and the dynamic playing field for IR, warrants further inspection. A brief summary of the development of the Internet and hyperlinks follows.

In an attempt to establish a United States lead in military science and technology, the US formed the Advanced Research Projects Agency (ARPA) within the Department of Defense (DoD) in 1957 (Kristula 2001). The requirement was set that it must be possible to maintain communications even in the event of a nuclear attack, and as a result the US DoD Advanced Research Projects Agency Network (ARPANet) was created as an experimental computer network twelve years later (KWMC 2001). ARPANet linked four nodes via 50 kilobits per second lines, and became the forerunner of the Internet as it is known today (Kristula 2001).

ARPANet continued to grow into the 1970s and 1980s, but was used almost exclusively for military purposes (MCPL 2001). As a result, the information stored by the linked computers was tightly controlled, and it was not possible for an outsider to publish personal information, or otherwise add freely to the information content of the Internet.

The first version of one of the most basic but most-often used programs on the Internet was written in 1972 by Ray Tomlinson: the e-mail program (Kristula 2001). At this stage a total of 23 nodes were connected. The availability of e-mail became one of the catalysts of free sharing of information between users – an activity which was unheard of during the ARPANet phase of the Internet.

One area where standardization on the Internet landscape was about to become firmly established, was that of hardware and networking protocols. Just after the establishment of e-mail, developmental work started on the network protocol which would become the basis on which the Internet was to be run – Transmission Control Protocol (TCP) (Kristula 2001). The term “Internet” was officially mentioned for the first time in a paper on Transmission Control Protocol by Vint Cerf and Bob Kahn in 1974 (Kristula 2001). The Ethernet network protocol, which would later become one of the most-often used network layers through which access to the Internet was made possible, was developed in 1976 by Metcalfe. A total of 111 host computers were connected to the Internet at this stage (Kristula 2001).

Another major software development, similar to the development of the first e-mail program, was the release of the first USENet program. Steve Bellovin, a graduate student at University of North Carolina, programmed USENet (a decentralized news group network). Next to e-mail, USENet was to become one of the most important initial reasons for users to access the Internet (Kristula 2001). However, as was the case with e-mail, once again there was no central body to control the content of discussions on USENet. Users could not only read the contributions of others freely, but also participate themselves without constraints or vetting of contents.

In parallel with ARPANet, the US National Science Foundation created a backbone for other networks called CSNET. It became obvious that a connection between these two networks would be advantageous, and a plan for an inter-network connection between CSNET and ARPANet was proposed in 1981. This connection implied further growth in content availability, since both networks would now contribute towards an increase in the amount of information available. At this point, 213 hosts were connected to ARPANet (Kristula 2001).

Further developments on hardware and protocol standardization were made in 1983, when the use of the Transmission Control Protocol – Internet Protocol (TCP/IP) for connection to ARPANet was prescribed. Up to date the mere fact that an Apple Macintosh user can send an e-mail message to an IBM PC-compatible user without having to do any conversion (completely different operating systems and hardware architectures are involved), is a positive result of the standardization on TCP/IP.

Another major technical innovation was the creation of the Domain Name System (DNS) by the University of Wisconsin. A domain is a unique address by which a computer on a network is known. Prior to the development of the DNS, such addressing had similarities to a long, cumbersome telephone number preceded by city and country codes. The DNS made it possible for a user to access another computer on this network by using a name instead of having to memorize any numbers. A website can be referred to either by its domain number, which is a set of four groups of digits (for example 196.2.63.110), or by its URL (for example <http://www.mweb.co.za>). Since both modern browsers (Microsoft Internet Explorer and Netscape Navigator) will complete the URL when a user omits the <http://> part, the URL method is clearly easier to use and remember than the number system. Compare for example, the ease of associating a URL with the contents of a website, and remembering the URL [www.bmw.com](http://www.bmw.com) as opposed to 192.109.190.103. A total of 562 hosts were now connected.

A major factor in the growth of the Internet, and one which was a large step towards access for the man in the street, was the creation of NSFNet by the National Science Foundation (1984). This network offered a way of connecting NSF computers to university computers to facilitate the sharing of research work. NSFNet, and other networks which evolved from it, eventually became known as the Internet (MCPL 2001). The NSF started using 1.544 Megabit per second communication lines, to cater for the increase in network traffic – a total of 1961 hosts were connected during 1985. This number jumped to 2308 during the next year, and increased more than twelve-fold to 28 174 one year later (Kristula 2001). This irregular mathematical progression in the number of connected computers did exhibit one clearly discernible pattern: continuous increase. Another approximate doubling in numbers occurred over the next year: from 28 174 to 56 000 hosts (Kristula 2001).

Strong emphasis on software development, including the WWW and hyperlinks, was evident in the growth of the Internet over the next few years. This development started with a major software breakthrough in 1989. A need was identified for a tool to allow collaboration between physicists and other researchers in the high energy community (Gribble 2001). One critical aspect was the need to transmit not only text, but also graphics. TCP/IP had already enabled the transmission of text, but these physicists needed the transmission of the accompanying graphs and other non-textual information. The authors of this tool, called the WWW, were Tim Berners-Lee and Robert Cailliau, from CERN in Geneva, Switzerland (CERN 2001). Their tool included a hypertext system, to assist researchers in sharing their information by using simple hyperlinks to other data (Kristula 2001, MCPL 2001). At the end of 1989 the number of hosts had increased to 313 000 (Kristula 2001).

The information-sharing system known as the WWW was based on the programming platform called HTML (Hypertext Markup Language), and the HTTP protocol and client computer software (known today as a browser program) (Gribble 2001).

The hypertext system was opened to public access in 1991, and the first commercial use of the Internet was permitted (Gillies and Cailliau 2000:319, MCPL 2001). The host total again increased to 617 000 (Kristula 2001).

The Internet Society was created in 1992 (Kristula 2001). Over 1 136 000 servers were established all over the globe by the end of this year (Gribble 2001). The next software breakthrough was motivated by the need of the user to navigate through the ever-growing mass of information available on the Internet. It came in the shape of the programming of a WWW browser called Mosaic during the next year. It was created by a group from the University of Illinois at Champaign-Urbana, led by an undergraduate student Marc Andreessen (Gribble 2001). This and subsequent Internet browsers played a big part in the explosive growth of the Internet user base (MCPL 2001). In parallel with the birth of Mosaic, InterNIC was formed to manage registrations, directory and database services (Kristula 2001). The number of hosts grew to 2 056 000 in 1993, again almost doubling to 3 864 000 one year later (Kristula 2001).

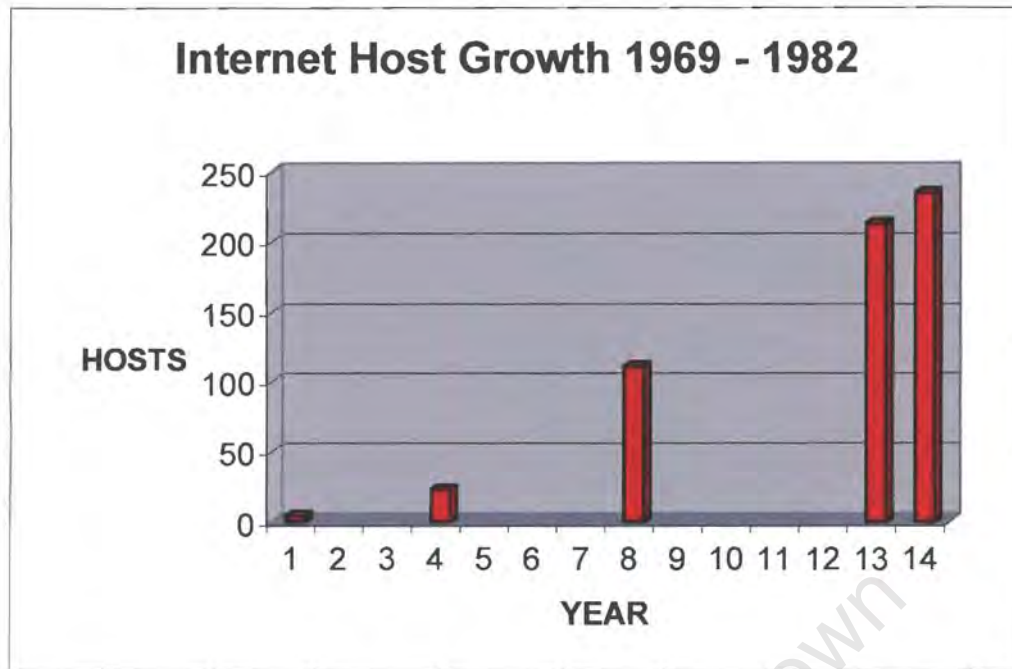
The explosive growth of the Internet, indicated by the approximate doubling in size every year, caused the NSF to prohibit direct access to its backbone in 1995. It contracted this function out to four other companies. A USD50 fee was imposed on registering domains other than .edu and .gov. This excessive degree of information availability is criticized by Chun quoting Stoll in 1995, who complained that "...the Internet is an ocean of unedited data, without any pretence of completeness... (it) has become a wasteland of unfiltered data" (Chun 1999:140).

The number of hosts again almost doubled to 6 642 000 in 1995 (Kristula, 2001). Hoffman, Kalsbeek and Novak estimated that 16.6 million users in the United States alone existed during this year (Hoffman, Kalsbeek and Novak 1995). Independent Internet Service Providers (Sprint, Uunet, MCI, AT & T, etc) started carrying most of the Internet traffic in 1996 – over 15 000 000 hosts were now connected (Kristula 2001).

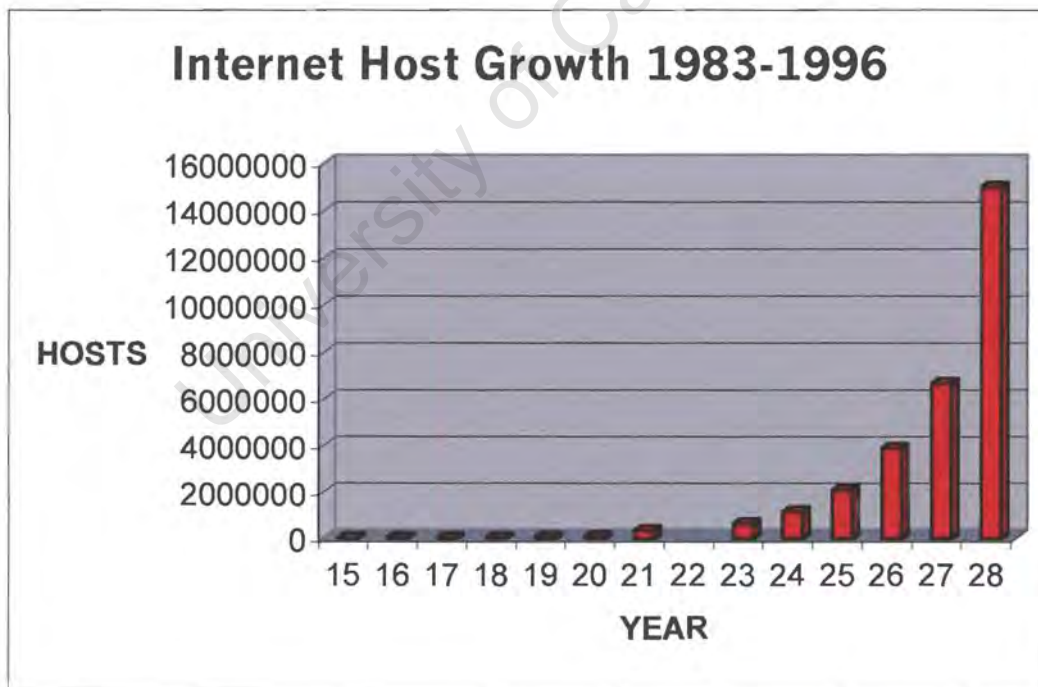
A summary depicting the exponential growth of the Internet, measured by counting the number of servers, follows in Table 2.1 and Graphs 2.1 and 2.2. For the sake of consistency, data from only one source was used for Table 2.1, resulting in some missing figures. Each year in this table was allocated a number, to make the graphs easier to interpret. Since the y-axis data in the graph spans such a large series of values (from four to 15 000 000), two graphs were drawn to minimise the effect of the lower numbers being displayed as a flat area instead of a visible bar.

YEAR	NO.	HOSTS
1969	1	4
1970	2	
1971	3	
1972	4	23
1973	5	
1974	6	
1975	7	
1976	8	111
1977	9	
1978	10	
1979	11	
1980	12	
1981	13	213
1982	14	235
1983	15	562
1984	16	1024
1985	17	1961
1986	18	2308
1987	19	28174
1988	20	56000
1989	21	313000
1990	22	
1991	23	617000
1992	24	1136000
1993	25	2056000
1994	26	3864000
1995	27	6642000
1996	28	15000000

**TABLE 2.1 – Growth of the number of Internet hosts**



GRAPH 2.1 – Internet Host growth Period 1



GRAPH 2.2 – Internet Host growth Period 2

The focus now moved away from the software and hardware developments to the Internet user and IR from the Internet. Figures of actual Internet usage as opposed to the number of servers also indicate exponential growth, and President Clinton stated in 1998 that 100 million users logged onto the Internet per day (Selwyn 1999:226). NUA claimed a USA/Canada total of 167.12 million users in November 2000, with a world total of 407.1 million (NUA 2001). Sherman echoed Stoll's sentiments by stating that the Internet had become the world's largest and most complex, chaotic and unstructured search space (Sherman 1999:54,57). Another author has found that retrieval and organization of materials on the web are not standardized, while others claimed that the WWW was not designed to support organized publishing of data, or the retrieval thereof (Coetzee 1999:10, Spink, Bateman and Jansen 1999:117). Information overload has arrived (Wyle 2001).

One method of looking for information on the WWW is to use a search engine. One researcher claimed that many practitioners consider Internet searching to be a very urgent, challenging and rewarding area for future research (Chen and others 1998:584). However, a study done to determine what portion of the web is covered by search engines, confirmed that it is a shockingly small percentage, varying between 3% and 34% (Sullivan 1999a:30).

The status quo of the Internet today clearly emerges from the picture painted above. It indicates an exponential rise in connected computers, implying a similar increase in the amount of information suddenly becoming readily available. An estimated 407.1 million users were online to the Internet by November 2000 (NUA 2001). It is this mass of unorganized information which the user faces, which is part of the focus of this research project.

#### 2.2.2.2 HISTORY OF HYPERLINKS

In a traditional library the user is limited in the content he/she can access, simply due to the physical and time constraints of not being able to move quickly from one book/magazine/journal to another relevant one.

This other relevant document probably does exist, but there is not necessarily a pointer to it and even if there is one, it is not always physically practical to continually retrieve the next document from the system and continue reading. However, the hyperlink concept changed this situation dramatically.

The landmark article by Bush in 1945 (mentioned in Section 2.2.1.2) identified a special point in a document, called an associative index. If a user performed a certain basic action at this point, such as pressing a keyboard key, he/she would then be transferred to another relevant point elsewhere (Bush 1945). Ted Nelson envisaged a similar situation with his Project Xanadu – he proposed a “docuverse” (document universe) where all writing is linked and referenced (Butner 2001, Ellis 1996:126).

Both these ideas are echoed in the Internet’s concept of hyperlinks – clicking on a hyperlink transfers the user to “another point”, which could be a different word/sentence/paragraph or picture on the same page, or a website halfway around the globe. Once again this freedom of “movement” contributed greatly to the freedom of content, which characterizes the information available on the Internet. Hyperlinks are features of the Internet, and as such provide an easy way to navigate the vast world of information.

### **2.2.3 IT IN EDUCATION**

Education in general is as old as mankind itself, but IT is a relatively young field. While there is agreement that these two areas can supplement one another with positive results on both sides, the age difference between the two resulted in this marriage experiencing an uncertain start. In parallel with the praise for and criticism of new methods used in education (as outlined below), some questions were being asked regularly in the literature, all referring to the role of IT in education:

- should IT be used in education?
- if yes, has IT been applied successfully in education?
- what is the effect of applying IT in education?

### 2.2.3.1 EDUCATOR VS LEARNER

The respective roles of the educator and the learner came under the spotlight when a change of emphasis was proposed in a number of studies. Ives and Jarvenpaa suggested that learners take more control of their own education, while the educator should change roles from teacher to facilitator (Ives and Jarvenpaa 1996:35). In support of this suggestion, Davis claimed that the teaching of the history of the technology would change the role of the student from being a passive receiver of knowledge (Davis 1996:852). Lagowski stated that the lecturer should no longer be the source of all information, but simply the facilitator to obtain it (quoted by Marais and Marais 1999:88). This study aims to provide a method for learners to find academic information on the Internet, which could assist in the shift of activity from the educator to the learner.

### 2.2.3.2 EDUCATIONAL METHODOLOGY AND TECHNOLOGY

A different issue emerging from the literature was the contrast between traditional and modern educational methods. Cookson listed the traditional methods as including face-to-face instruction, printed matter, radio and television broadcasts and audio/video recordings (Cookson 2000:72). The same author implied that higher education had been placed in a mould of outdated techniques as long as seventy years ago (Cookson 2000:79). Marais and Marais implicated a specific educational field in an attack on this traditional educational methodology.

They claimed that education in the chemical engineering field (in most institutions) for example, had remained stagnant for the last 100 years, and students experienced a passive environment where there was one-way interaction from educator to learner (Marais and Marais 1999:82,87). One US college student was quoted by the same authors as having his/her hopes for learning and intellectual debate dashed as parrot-fashion regurgitation of knowledge and simply making certain grades were the only requirements from the educators (Marais and Marais 1999:88).

Ives and Jarvenpaa suggested that a similar situation existed in IS education, by referring to "... memorisation..." and the absence of direct involvement of the learner with the real world (Ives and Jarvenpaa 1996:35). Furthermore, Malan stated that the overheads of a given institution could only support face-to-face instruction as the main mode of delivery up to a certain point (Malan 1999:60). In the same vein, other authors urged educators to rethink their approach to education (Quilling, Erwin and Petkova 1999:5).

On the other hand, modern educational methods were broadly defined as being knowledge media: computing, the Internet and communications (Cookson 2000:72). As far back as the early 1990s at least eight references were made by Corritore and others to computer laboratory components having been discussed in the literature (Corritore and others 1999:292). Off-campus teaching was also recognised as a growing trend in education (Cookson 2000:72, Tearle, Dillon and Davis 1999:14). Selwyn quoted Maddux, who believed that the Internet had the potential to become a highly significant educational tool (Selwyn 1999:225). Marais and Marais also believe that the acquisition of skills was more important than the passing of examinations, and that technology-based teaching (Internet, e-mail, cable and satellite broadcasting) should play a bigger role (Marais and Marais 1999:88).

On the one side of the "for-against" argument, a number of experiences were noted where IT had been applied in education with distinctly positive results. Based on four years of classroom experience, Edling claimed that IT enhanced student learning, reduced the time taken to teach and allowed deeper coverage of materials (Edling 2000:10). This author also decreased traditional lecture time almost completely in favour of students using computers on their own, with enhanced student learning as result (Edling 2000:10,13). Cookson quoted Bates, who listed the following eight important positive results of using the Internet in higher education:

- it expanded learner access to training and education,
- it caused an increase in education quality,
- it caused a decrease in the cost of education,

- it raised the cost-effectiveness of education,
- it increased the number of courses on offer,
- it generated higher income,
- it enabled the generation of specialised study programmes, and
- technological innovation had a positive influence on other aspects of the operations (Cookson 2000:73).

In the centre of this argument, some authors were cautious about the use of IT in education. Wallace, Kupperman and Krajcik admitted that research on the use of IT in the classroom was in its infancy (Wallace, Kupperman and Krajcik 2000:75). Mellon spoke out against falsely upholding technology as the solution to all educational problems, and believed that past predictions of books being replaced by technological media had been proven incorrect so far (Mellon 1999:28,29). Gan stated that IT as used in Malaysian education circles should be applied with care (Gan 1998:32).

On the other side of the argument, some downright negative experiences were recorded. Selwyn claimed that the use of IT in education for the last 20 years had had consistently disappointing results (Selwyn 1999:227). This author believes that the installation of computers or connecting them to the Internet was often economically and not educationally motivated. In the UK specifically, Internet access in British schools (late 1997) was much less prevalent than what was planned, with more than half of existing computers incapable of running Microsoft Windows (Selwyn 1999:227).

A number of claims about the reasons for educational failure were made with regard to Information Science education in developing countries. Saracevic stated that a shortage of information professionals was the number one problem in education (Saracevic, Braga and Afolayan 1984:192). Further shortcomings included a lack of clear direction in Information Science, technology problems and a low level of basic computer literacy with information professionals (Saracevic, Braga and Afolayan 1984:194,196).

These authors also claimed that although computers were "... common in most developing countries... ", education about their use, especially as a tool in information services, was not (Saracevic, Braga and Afolayan 1984:196). This research project sets out to inspect the use of such a tool, namely search engines, and whether or not intervention in the use of these tools is necessary to ensure positive results.

Cookson summarized the above two contrasting points accurately (the success versus the failure of using IT in education) by stating that "... learning environments created by the unfolding digital technologies pose both opportunities and threats to higher education institutions" (Cookson 2000:71).

It becomes clear from the above that the current status of IT in education is very volatile. There is reason to believe that modern methods could be more successful than traditional "talk-and-chalk" methods. At the same time, however, throwing technology at a student without careful planning and implementation creates new problems without solving the old ones. The author does believe that IT should be used to enhance education, but at the same time there have been some definite failures in its implementation. The effects of applying IT in education are difficult to quantify, and an attempt will be made with this research project to measure this effect in one specific area of application as accurately as possible.

## **2.3            THE LIBRARY**

### **2.3.1            THE ROLE OF THE WWW**

It is evident from the literature that the role of the library and the librarian in education is changing, resulting from the influence of technology. The title of a collection of essays by White underlines this fact: "At the Crossroads: Librarians on the Information Superhighway" (White 1995). The use of a library has always been fundamental to any study, research and education, and will probably remain so.

Marcella claims that more than three-quarters of the respondents to a mailed survey on information needs in the UK claimed that they would use a public library at least occasionally to satisfy their information needs (Marcella and Baxter 1999:175).

However, rapid advancements in the IT/IS field have made it difficult for libraries to keep up with the constant demand for new texts in these fields. Budget constraints and updating logistics compounded this problem. It was a natural move to merge the function of a library and that of the Internet as resource providers. Some authors viewed the incorporation of the Internet into the service offered by a library as a challenge (Radcliff, Du Mont and Gatten 1993:17). Others found that the largest application for hypertext systems in academic libraries in the UK was for networked document retrieval (Furner-Hines and Willett 1995:31). This indicated that this merging had been taking place. It would however be naïve to expect that a total merge between the traditional IT department and the library should take place at all institutions of higher education. IT departments normally serve other sections as well: the academics as vehicle in education, and the administration as support for registration and financial matters, to name but two. In this situation, the Internet can be seen as an enabling technology rather than as a choice to be made.

Voorbij claims that libraries should support their users by assisting them in searching for information on the WWW (Voorbij 1999:598). However, it was found that librarians, traditionally a certain source of answers for the information-seeking learner, experienced problems in dealing with the Internet as medium. Subject librarians did not always find what they needed using the searching facilities on the Internet. Basu has shown that 29.6% of a study group of reference librarians found the information they needed between 21% and 30% of the time, while only 14.8% of them found it more than 50% of the time (Basu 1995:38). Problems associated with the use of this method to find information included lack of time and the quality and quantity of data returned (Basu 1995:38,39). Another study showed that some librarians were unable to supply answers to the questions put to them owing to a lack of sophistication on their side (Devlin and Burke 1997:104).

Most authors referred to the paradigm shift: librarians had to start acquiring new competencies to be able to survive (O'Leary 2000:21,22), and information professionals had to pro-actively search the web to filter out unwanted information for the users (Burton 1999:104). Corcoran, Dagar and Stratigos claim that "...traditional business models of library operation are defunct...", and that although the number of requests for information had decreased, the complexity level had increased, implying that the expertise level of the service provider must also rise (Corcoran, Dagar and Stratigos 2000:29,30). Sherman quoted Kahle, who states that for every one individual entering a library per day, 20 web searches are being done, implying that the web is beginning to challenge the value of all libraries put together (Sherman 1999:61). Two authors describe in detail how three high-profile libraries (Los Alamos National Laboratory, Shell Research and Hewlett-Packard) have adapted dramatically to be able to deliver a cutting-edge service (Pack and Pemberton 1999a, Pack and Pemberton 1999b; Pemberton and Pack 1999).

### **2.3.2 THE WWW AS SOURCE OF EDUCATIONAL INFORMATION**

It can safely be stated that the Internet provides a rich pool of information to any user. Large and Beheshti states that the WWW is not an information source, but rather a link to millions of other sources (Large and Beheshti 2000:1069). Some facts can however not automatically be assumed to be true, and need to be addressed here:

- is there educational information available on the Internet?
- if yes, can learners find this information?
- does learning take place during this process?

The first question above is addressed by Bar-Ilan in an attempt to determine if the WWW can be used as an information resource for scientific research. In conclusion it is claimed that "... valuable, freely available data is hidden in the Web waiting to be extracted..." (Bar-Ilan 2000:432).

The second question is of too wide a nature to have a specific answer. A number of studies have been done on searcher behaviour and information seeking success, all involving school children or students searching for subject materials on the Internet (Fidel and others 1999, Hirsh 1999, Large and Beheshti 2000, Oliver and Oliver 1997, Spink and Saracevic 1997, and Sutcliffe, Ennis and Watkinson 2000). The degree of success achieved varies widely, and depends on a large number of factors for some of the studies. However, this project will attempt to answer this question by involving learners in searching experiments on a large scale.

A study by Oliver and Oliver addresses the third question above. A survey was done with schoolchildren to determine if learning automatically takes place during an information-seeking task, and found that it does not (Oliver and Oliver 1997:519). From previous studies by these authors it was clear that one possible reason for this failure could be that participants concentrate more on the mechanics of the seeking process than on the resultant information. However, this study apparently only involved information seeking on static data bases such as CD-ROMs, and excluded the Internet as source.

Another study done by Voorbij has shown that many users do experience problems in finding subject-specific information on the Internet. Only one-third stated conclusively that they found it easy to locate subject information (Voorbij 1999:604).

During an extensive study on the searching behaviour of high school learners, Fidel found that the participants "...performed focused searching..." and managed to progress through the searching process at a high rate. It was concluded that the WWW has tremendous potential as a tool for information gathering (Fidel and others 1999:36).

Hirsh did a similar study on elementary school children, and found that they "... are increasingly using electronic resources to support information needs..." (Hirsh 1999:1265).

However, there is growing concern about the difficulty experienced by learners, educators and others in accessing educational resources on the WWW (Sutton 1999:1182). A study by Large and Beheshti identified frustration among many primary school children when the exact information sought was not found on the Internet (Large and Beheshti 2000:1079). At this stage it is not clear what the situation is with learners in higher education institutions. In this project the author will attempt to determine this status quo.

## **2.4 CASE STUDIES AND EXPERIMENTS**

Applebee and others did a general case study on use of the Internet by Australian academics. Some of the findings were:

- there is a need for training on the use of the WWW,
- most academics use the WWW regularly,
- time constraints prevent academics from exploring the web or being trained on its use, and
- it is possible that students are far more familiar with Internet usage than their professors (Applebee and others 2000).

A variety of other case studies on the use of IT in general and the Internet specifically were done in a number of fields from medical science, primary school teaching to higher education.

### **2.4.1 INFORMATION RETRIEVAL**

The IR-related works were classified into three groups according to each one's context: schoolchildren, the medical field and real-life searching situations.

Firstly, Wallace, Kupperman and Krajcik did a six-week empirical study on sixth-graders in a science class. They found that emphasis should not be put on the technology, to prevent the attention from being focussed on the computer. It was also not possible to transfer information seeking skills to the learners (Wallace, Kupperman and Krajcik 2000:75). It appeared that successful searching involved a decrease of the focus on technology.

These results served as a pointer for this research project, since the following specific facts indicated a broad direction to follow:

- although participants seemed to enjoy the online experience, they did not seek to understand what they were doing,
- the information seeking process was treated as a separate activity to that of learning, placing the focus on the technology, and
- little progress was made in convincing the participants to engage in the complex cognitive strategies required for IR (Wallace, Kupperman and Krajcik 2000:100).

The author decided to take particular note of these findings while designing the proposed model.

The second group involved IR studies in the medical field. Hersh, Pentecost and Hickam did a study on medical students' usage of two medical searching systems, where one was based on Boolean operators, and the other on natural language searching. The results proved that both types could be used effectively with minimal training (Hersh, Pentecost and Hickam 1996). Again it appears as if the technology on which the searching process is based (in this case the actual operators used) does not play an important role in searching success. The author would also include this finding in the design of the proposed model.

Another study investigated methods to improve access to information for remotely stationed medical staff. It identified a need for nurses working on the remote Western Isles of Scotland to access sources of information needed in individual patient situations. Other research, however, found that medical staff in general did not make use of information sources (Farmer, Richardson and Lawton 1999:60).

Howcroft and Mitev did an empirical study on the use of the Internet in medical practice management in the UK. It was found that if doctors had the time to search for information, they would use it to find relevant information.

Further phases of the study were halted since funding had been withdrawn (Howcroft and Mitev 2000). In both of the last two cases, it appears as if a need does exist for medical personnel to access online information. However, financial and time constraints are impeding the shift from traditional to online sources for users in the medical field.

The last group includes studies done using real-life situations, including searching intermediaries. Saracevic did an early detailed study on users searching for information. Experiments were done under real life situations, and involved the user context of questions, question structure, decision-making by searchers and how the same question could be searched for in different ways (Saracevic and Kantor 1988a, Saracevic and Kantor 1988b, Saracevic and others 1988). However, this study was done before the availability of the WWW and hyperlinks, and as such is useful for general IR from electronic sources and not IR from the Internet specifically.

Spink and Saracevic studied the sources and effectiveness of search terms used during mediated on-line searching under real life circumstances. The interaction between 40 users and four professional intermediaries was analysed using video tapes, transaction logs and dialogue (Spink and Saracevic 1997:741). It was found that the search terms contained in the users' written question statements produced the most relevant retrieved items. The process of selecting search terms and the construction of a search query were found to be highly interactive and a non-linear dialogue (Spink and Saracevic 1997:760).

Barry attempted to determine the extent to which various document representations contained clues that allow users to determine the presence, or absence, of qualities that determine the relevance of the document to the user's situation. It was found that the way a document is presented is not an effective indicator of potential relevance. Different types of document presentations have different abilities to supply clues to the searcher on various qualities of the document (Barry 1998:1293).

Khan and Locatis did two studies to determine the effect of link density, link cues, link correspondence and display format on search performance. It was found that experts performed better than novices, since they were better able to prioritize tasks (Khan and Locatis 1998b:176). It was also found that a correspondence of wording in search tasks and wording in links had a positive effect on search efficiency, but not on search accuracy (Khan and Locatis 1998a:1248).

The Voorbij study showed that many users do experience problems in retrieving subject-specific information from the Internet. Only one-third stated conclusively that they found it easy to locate subject information (Voorbij 1999:604).

An up-to-date study done by Spink and others, confirmed that users make little use of advanced search features. For example, a test on 1 025 910 queries submitted to the Excite search engine revealed that the following features were only used by small percentages of users, as indicated:

-	Inclusion (+)	2%
-	Exclusion (-)	0.001%
-	Boolean operators (NOT / AND NOT)	0.0003%

(Spink and others 2001:229). Another study seemed to indicate that searchers spent a relatively short amount of time searching for one topic: the average search session seemed to last between five and 10 minutes only (Cooper 2001: 137). This could possibly be a reason for the lack of use of advanced operators.

#### **2.4.2 SEARCHER BEHAVIOUR**

Marchionini defines information seeking as "... a set of behaviours controlled by interacting levels of conscious (logical) and unconscious (intuitive) mental activity" (Marchionini 1995:62). The prior reading on research work, discussed here, focuses on this set of behaviours.

However, Large, Tedd and Hartley warn that information seekers must not be treated as a homogenous group – they differ in many aspects, where their IR experience level has a large differentiating effect (Large, Tedd and Hartley 1999:29). Furthermore, Jansen and Pooch warn that there is a difference in the way that WWW searchers and traditional OPAC searchers work when looking for information (Jansen and Pooch 2001:244). In contrast to this finding, a remarkable similarity was found between the basic methodology of early online searching and modern Internet searching. Lancaster listed a four point sequence of steps to be followed for online retrieval: logging on, negotiating the search process, manipulating the results and logging off (Lancaster 1978:72). Some of the older studies involved searching an electronic source, excluding the Internet, and this factor will be used below to distinguish between two groups of studies.

#### 2.4.2.1 NON-INTERNET SEARCHING

An early study by Martin on the users of the ESA-IRS system indicated a worrying trend: many searchers appeared to ignore the interactive capabilities of a searching system, and seemed to use only the most basic features of the feature set at their disposal. Sophisticated command functions were almost totally ignored, and it was noted that “occasional users” form a larger part of the searcher community than what was believed at this point. This group of users appear to have special requirements in terms of program interfaces, and this author suggested that the IRS be approached to address these special needs (Martin 1982:67).

The effects of subject knowledge and previous search experience on the use of online search tactics by both novices and experienced users were inspected in a study by Hsieh-Yee. Although the study did not include Internet searching, it nevertheless provides useful insight into searcher behaviour. The results proved that searcher’s use of search tactics was affected by search experience. It also proved that only after a searcher had gained a certain amount of experience in searching, subject knowledge or lack thereof became a factor (Hsieh-Yee 1993:161).

A study was done to see how advanced humanities scholars as end users did searching on online databases. The use of operators was included in this study, where some of these were also available with Internet search engines (Boolean and proximity operators). A fairly high percentage of participants used a simple single word as search term, and used no Boolean operators at all. In both cases this indicated a lack of searching sophistication (Siegfried, Bates and Wilde 1993:273). The author realized that if this finding was echoed in the results of this project, it would require specific measures to be taken to solve the problem.

Taylor's definition of the four stages of query formulation held promise of assisting in the design of the proposed model. The difference between the visceral and the compromised information need of the user should be mentioned here (Ingwersen 1992:112,113). Since the model is aimed at the average user, it could not be assumed in the design that the users would be experts on search engine syntax, advanced searching concepts and effective query formulation. Therefore this difference was bound to be larger than the difference for an experienced intermediary searcher, for example.

Yuan attempted to determine the relation between end-user search experience and searching behaviour. The purpose was to gain a better understanding of end-user searching behaviour to allow search system designers and trainers to provide better systems and programs for users (Yuan 1997:218). Although it was found that search experience does affect some aspects of end-user searching, no general conclusions in this regard could be made (Yuan 1997:231).

Sutcliffe, Ennis and Watkinson did an empirical investigation of IR using the MEDLINE database to study user behaviour and performance (Sutcliffe, Ennis and Watkinson 2000:1211). The overall performance of searchers was poor, although poor performers often displayed many qualities of good searchers, but failed for only one reason. The behaviour of searchers is complex, since sub-optimal strategies can compensate for other difficulties in searching displayed by the same searcher performance (Sutcliffe, Ennis and Watkinson 2000:1228).

It does appear as if searching electronic sources for specific information is not an easy task. Searching behaviour is a complex topic, with many unanswered questions.

#### 2.4.2.2 INTERNET SEARCHING

A study done on hypertext searching has proved that there is no significant relationship between the searcher's confidence and the number of pages visited (Wolfram and Dimitroff 1997:1144). Another study by Ellis, Furner and Willett on hypertext links found that no conclusion could be drawn about the relation between consistency and effectiveness during the assignment of index terms (Ellis, Furner and Willett 1996:287,298). However, it was found that confident searchers spent much less time searching than those with less confidence (Wolfram and Dimitroff 1997:1145).

Fidel and others did a study to analyse the Web searching behaviour of high school students while searching for information for homework assignments. The participants' discussions as they were busy searching were audiotaped and used as later evidence of findings (Fidel and others 1999:25). It was found that searchers were satisfied with their results in general, but complained about slow responses (Fidel and others 1999:24).

A study was done by Hirsh to explore the relevance criteria and search strategies applied by elementary school children while searching for information. It was found that the searchers displayed little concern for the authenticity of the materials found, and spent a large amount of time on finding pictures (Hirsh 1999:1265). Most searchers also could not formulate effective search queries; they would misspell search terms and did not use Boolean operators (Hirsh 1999:1268). This lack of concern about the validity of resources was echoed by the findings of Grimes and Boening. They found that learners used unevaluated web resources, and that instructor expectations and actual learner usage of web resources differed widely (Grimes and Boening 2001:11).

Marcella and Baxter did a study on the information needs and information seeking behaviour in the UK (Marcella and Baxter 1999:175). This study involved a large number of electronic sources, including the Internet. Only a small percentage of the respondents preferred the use of a computer above traditional methods (books, libraries, friends, etc) to satisfy their information needs, while a large percentage indicated libraries as preferred source (Marcella and Baxter 1999:172).

Large and Beheshti interviewed 50 schoolchildren to find out how they searched for information on the Web (Large and Beheshti 2000:1069). Sophistication in the IR strategies of these children was evident, admittedly after having been shown at least two basic search methods (Large and Beheshti 2000:1076). This finding was to be considered in the design of the proposed model for this project – a short training session on the use of the model would be built into the searching experiments to be done.

Ross did an empirical investigation of the subject content of queries submitted to the Excite search engine (Ross and Wolfram 2000:949). Since three or even four dimensions were necessary to map relationships between topic contents, it became clear that the inter-topic similarity required a complex set of dimensions to describe it (Ross and Wolfram 2000:957).

Again (as with Non-Internet searching of Section 2.4.2.1), there appears to be a pattern of many searchers experiencing failure in their endeavours, while some did have success. Either way, no reports of completely trouble-free searching were found, and at this stage it was clear that assistance in whatever form would be welcome.

## **2.5            SEARCH ENGINES**

### **2.5.1        INTRODUCTION**

Lancaster predicted a change in the searching process when he described a future scenario in 1978:

"The scientist of the year 2000 will use a terminal in many different ways: to receive, transmit, compose and search for text, seek answers to factual questions, build files and converse with colleagues. The terminal provides a single entry to a wide range of capabilities that will substitute, wholly or in part, for many activities now handled in different ways" (Lancaster 1978:331).

Lancaster's "scientist" is today's average PC user, his "terminal" has become the desktop computer, "conversing" has evolved into e-mail and his "different ways" could be the shift from using an intermediate searching expert to doing the search oneself.

Search engines provide the average Internet user with a free, apparently easy way to find general information on the Internet. They are programs which offer interaction with the Internet through a front end, where the user can (at the least) type in a search term, or make successive selections of relevant directories. The search engine software then compares the search term against an index file, which contains information about many websites. Matches found are returned to the user via the front end. The index is updated regularly either by human editors or by automated programs (called spiders, robots or crawlers). Both humans and spiders simply collect information of new websites by visiting as many websites as possible, and then building them into the index. The three components of a typical search engine (front end, index file and information collectors) have close parallels in the components of a typical IR system, as defined by Lancaster many years before the Internet and search engines became freely accessible. A search engine front end maps to the "user system interface", the index file maps to the "indexing subsystem" and the information collectors map to the "document selection subsystem" (Lancaster 1978:13). Another parallel is evident between the information collectors and this author's description of the functions of the human analysts who did the manual indexing of the MEDLARS system. A six-point list of their duties closely matches the tasks of a human editor building a search engine database (Lancaster 1978:39).

This author also refers to the importance (in early IR systems) of choosing the correct data base, the user's insight into its use and his/her ability to match the information need to his/her technical searching abilities (Lancaster 1978:154). This reference to and guidance about the data base holds remarkably true for modern day search engines and their indices.

During 1992 an alphabetic subject listing of WWW resources appeared on the opening CERN page. The first Internet search engines as they are known today started appearing in 1994 (Schwartz 1998:974). In 1995 Basili warned that there was an urgent need for standardized tools to find, and describe the information available on the Internet (Basili 1995:459).

Schwartz also claimed that in 1996 search engine programs started featuring on television, newspapers and in trade journals. At this time a difference in the operation of products became evident: it was possible to distinguish between directories, meta search engines, subject-specific directories and personal desktop searching agents (Schwartz 1998:974).

During 1997 it was claimed that approximately 200 search engines existed on the Internet (Gittlen 1997:17). Search engines were referred to as programs which made it easy for the user to find relevant data, but their real value would only be known with time (Dragan and Boscardin 1997:192).

Schwartz stated that there could be a move towards smaller but better resources operating in a richer environment in the future (Schwartz 1998:979), while Feldman claimed that the quality of results appeared to increase from one year to the next (Feldman 1998). However, a claim is made that search engines are complex, trusted without being understood and that users simply deal with their answers without understanding why they receive those answers (Lynch 2001:17).

## 2.5.2 INTERFACE DESIGN

As noted above (see Section 2.5.1), the front end of a search engine is the interface, the component through which the user specifies his/her search query, and through which answers are displayed. As with any computer program, this component plays an important part in the success the user has with the use of the program.

One author lists eight general interface design principles, which should satisfy the user and provide a high chance of success:

- consistency of design,
- availability of shortcuts for the experienced user,
- informative feedback,
- clear sequence of actions, with indication of completion,
- clear, simple error messages,
- action reversal should be possible,
- user must be in control, system responses should be visible, and
- minimal demands should be put on a user's short term memory (Large, Tedd and Hartley 1999:200).

The design of search engine interfaces was criticised by Large and Beheshti , after having done some searching experiments with 50 grade six schoolchildren. They believed that long lists of irrelevant answers could be suppressed to enhance a child's experience of searching the WWW for relevant information (Large and Beheshti 2000:1079).

Thomas did a detailed study on the effect of interface design on the selection of items in an online catalogue. This set of empirical experiments involved four types of displays, all simulating online catalogues. The displays varied in the amount of data displayed. During the first one of two tasks, participants appeared to be able to make better selections from screens containing less information. However, during the second task participants were able to make relevance judgements more often when using enhanced-content interfaces (Thomas 2001:20).

It was found that brief catalogue displays should be redesigned to include more fields with subject-rich content, and that only three fields seemed to be favoured for relevance judgements (Thomas 2001:42-43). Bilal criticized the design of the interface to the Yahoooligans! web search engine ( a product aimed at children), by claiming that it did not model children's information-seeking behaviour (Bilal 2001:134).

The author's personal experience with search engines confirms that interface design appears to have been neglected. Some examples from real life follow:

- Hotbot – text small, difficult to read, cluttered layout (Hotbot 2001)
- Opentext – search box difficult to find, objective of program hazy (Opentext 2001)
- 37.com – black background confusing, does not print well (37.com 2001)

### 2.5.3 OPERATORS

The syntax of search engines differs widely, and a wide variety of operators exists to allow focusing of a search query. Lancaster referred to natural language searching, where no operators are used, in an early reference (Lancaster 1978:279). Some of the more commonly used operators include phrases (using quotes), inclusion operator (+) and exclusion operator (-). Other, more advanced and less often used ones include Boolean operators (AND, OR, NOT), proximity operators (NEAR), stemming (\*) and field limiters (DOMAIN, TITLE). These operators are discussed by a variety of authors (Boulton 2001, Habib and Balliot 1999, Hock 1999, Notess 1997, Notess 1999b and Sullivan 1999a).

Boolean operators are commonly offered as search operators not only by most Internet search engines, but also by other electronic search services. These operators stem from the work of George Boole, a mathematician and logician who founded Boolean logic and algebra in the mid 1800s. One of his seminal papers was: "An Investigation of the Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities"

This concept of logic relies on the fact that any complex system containing variables with only two possible values, can be simplified by applying some basic rules (Voss 2001). As applied in searching, a typical Boolean search query could be:

1995 AND results AND ("world cup") AND rugby NOT (football OR soccer OR cricket)

Lancaster used similar Boolean searching examples in the MEDLARS searching system (Lancaster 1978:53,57). However, owing to the lack of modern technology, a simple Boolean search with three variables involved two successive searches, using slow sequential magnetic tape to store interim results (Lancaster 1978:61).

Frants and other authors claim that Boolean searching features in most retrieval systems (Frants and others 1999:86). These authors have shown that criticism of Boolean systems is mostly directed at the methodology employed rather than the Boolean principle itself. However, they regard Boolean and other search principles as being of equal importance (Frants and others 1999:94). A full description of the usage of Boolean operators, both in general databases and Internet search engines, is given in a book on electronic information seeking (Large, Tedd and Hartley 1999:144-152).

Hirsh found that during information seeking on electronic resources, children generally either do not use any Boolean operators or experience problems when they do (Hirsh 1999:1266,1271). During a 1996 task-oriented assessment using two medical searching systems, a traditional Boolean system was compared to a natural language system. No significant differences were found in the success rate of the two, as minimal training was given in both cases (Hersh, Pentecost and Hickam 1996:50).

Jansen and Pooch quote a number of other studies who claim a low percentage of real life searches having included Boolean operators: the Fireball study (less than 3%), the Webcrawler study (12%), Infoseek (10%) and Excite (29%) (Jansen and Pooch 2001:237,238,239).

A critical analysis of search engine accuracy, search functions and operational logic has proved that search operators are not 100% reliable. Experiments with the "link" and "host" limiting operators gave inconsistent answers (Snyder and Rosenbaum 1999:380).

In summary, a recent study claimed to have found that "... most people use few search terms, few modified queries, view few Web pages, and rarely use advanced search features" (Spink and others 2001:226).

#### **2.5.4 QUERY FORMULATION**

In general, the best way to construct a search query during an Internet search is not obvious to the average searcher. A common approach of these searchers is to simply type in one or more consecutive words into the search box. Where some search engines would add an implied OR operator between words (Infoseek), others will insert an AND (AltaVista and Google in their basic searching modes), and some might even treat it as a phrase (Looksmart). These three widely differing approaches would produce very different results, which would confuse rather than enlighten an average searcher. As an example, an Internet search using five words with AND operators between them will produce  $x$  answers from a given database. The same search using the same five terms and the same database, but using OR operators between the words, will produce  $y$  answers, where  $x \ll y$ .

It became clear from the literature that many users found the formulation of a search query difficult. Lancaster proved as far back as 1968 that the low quality of query formulations was the main reason for the failure of a search (Frants and Shapiro 1991:16).

This was also evident even in searching via an intermediary on non-web electronic database sources in 1993 (Su 1993:97). Hirsh found in 1999 that elementary school children "... have difficulty formulating and modifying search queries ... and have difficulties with spelling and Boolean logic" (Hirsh 1999:1266).

During a survey of 316 Excite users, Spink, Bateman and Jansen stated that few users employed logical operators and even fewer used the syntax correctly. They also had problems with search phrases and with the construction of "good search terms" and "complex search queries" (Spink, Bateman and Jansen 1999:123,125). Wallace, Kupperman and Krajcik had similar experiences in a study where sixth-graders had ample time to search for information – they found it problematic to convert their question into a query (Wallace, Kupperman and Krajcik 2000:87). Bilal also confirmed this phenomenon by referring to the "... naïve Web navigational skills..." of a group of seventh grade science students who used the Yahoooligans! web search engine (Bilal 2001:134).

A possible explanation for this phenomenon is the fact that the user had a question to which the answer was not yet known, and as a result did not know how to formulate the query (Mizzaro 1998:307).

However, Kassler claimed that the Internet could be mined for information which could provide a competitive advantage, if attention was given to search strategy and proper use of the available features (Kassler 1997:45). A number of attempts were made to rectify this situation by some authors who offered an alternative to the searcher constructing his/her own query. An algorithm was built which automatically constructed a Boolean query by Frants and Shapiro. This algorithm showed promise for successful implementation, since it accepted user feedback which was used to correct query formulations (Frants and Shapiro 1991:16).

Gauch and Smith tested an expert system for online search assistance, which automatically reformulated queries in an attempt to improve search results.

This system decreased the number of queries required, it increased precision and improved the relevance ranking (Gauch and Smith 1993:124).

The apparent problems searchers experienced with searching the Internet in general, and with query formulation and syntax specifically, gave reason for concern. The author decided to ensure that these problems are addressed in the proposed model to be built for searcher assistance.

### **2.5.5 SEARCH ENGINE EVALUATION**

Many authors have done evaluations, comparisons, measurements and a variety of other tests on a large number of search engines. These include: Chun 1999, Clarke and Willett 1997, Collins 1997, Conte 1996, Courtois and Berry 1999, Dong and Su 1997, Dwyer 1997, Garman 1999, Hock 1999, Hoff 1997, Huwe 1999, Lidsky and Kwon 1997, Nobles 1997, Notess 1995a, Notess 1995b, Page 1997, Sherman 2000, Singh and Lidsky 1996, Taubes 1995, Venditto 1996, Winship 1995 and others.

In these studies, the following search engines were mentioned:

AltaVista, Excite, Galaxy, Harvest, Hotbot, Infoseek, Looksmart, Lycos, Magellan, MetaCrawler, Northern Light, Open Text, PlanetSearch, Savvysearch, Search.com, UKPlus, Webcrawler, WWW Worm and Yahoo!. A list of commonly used search engines and meta search engines with their URLs and a short evaluation, is available on the WWW (Weideman 2000b).

However, in the studies mentioned above, a variety of methods were being used, and it is impossible to compare the results of the different evaluations with each other. Oppenheim and others did a study on this issue, and identified the difficulty of establishing recall on the fast-changing Internet scene as being one of the major reasons why different evaluations were difficult to compare. The creation of a standard set of tools will solve this problem and will also allow the tracking of the variation in performance of a search engine over time (Oppenheim and others 2000:190). The problem of measuring recall could remain unless the search engines are compared on closed data bases where prior evaluations of pertinence of the contents have already been made.

Landoni and Bell also underlined the importance of a scientifically sound approach to search engine evaluation. A number of traditional evaluation criteria were listed, including precision, recall and relevance ranking. It was advised that the web community should collaborate on finding a solution to the problem of search engine evaluation (Landoni and Bell 2000:124,126,129).

A study by Nicholson showed that replication could transform unreliable results to reliable ones by multiple iterations (Nicholson 2000:728). Various search engine categories were defined in different ways in an attempt to group programs, which are similar in function. Cohen and Jacobson distinguished between a subject directory and a search engine (Cohen and Jacobson 1999:1), while others classified a meta search engine as being different from a standard search engine (Garman 1999:75, Mickey 1999:80). A standard search engine has its own (large) data base, and a query submitted by a searcher via the search engine interface (front end) is compared to the content of this data base. The answer to the query is again communicated to the user via the search engine interface. A meta search engine, however, normally has neither its own data base nor the mechanisms to develop and maintain it. Instead the interface converts the user's search query into a form acceptable to a number of other, existing (standard) search engines, and submits it to these search engines in parallel. Upon receipt of the answers from the data bases of these sources, the meta search engine will collate the answers, and present it to the user as a single list.

Some severe criticism was given of a variety of inconsistencies of search engines by Snyder and Rosenbaum after a 1999 analysis. Firstly, a question mark was put over the reliability of results from search engines. Multiple searches on the same database have proved that entire categories have been ranked differently. Secondly, it was claimed that the functionality of search engines did not allow detailed web analysis. Both these shortcomings were believed to be market related rather than inherent weaknesses of the web-searching technologies (Snyder and Rosenbaum 1999:381).

A fresh approach to the measurement of the quality of answers produced by a search engine was taken by Losee and Paris in a 1999 study. Normally, measures such as recall and precision are used to evaluate a search engine. However, these authors considered the following two factors: the probability that the search engine is optimal, and, for a given query, the difficulty associated with retrieving documents. Their results have proved that this measurement method could give a more accurate indication of search engine success than more traditional measures (Losee and Paris 1999:882,888).

An empirical study was done on the level of precision of the first 20 answers returned for each one of 15 different queries on five well-known search engines. All five search engines performed acceptably in a variety of categories, including short, unstructured and structured queries respectively (Leighton and Srivastava 1999:870).

It was proved conclusively that the WWW does contain valuable research information, and can be tapped as a source of free data (Bar-Ilan 2000:432). This information was retrieved using standard search engines, all on the same date. Limitations that were noted included the fact that current search engines only allowed free text searches (Bar-Ilan 2000:441).

It is clear that the evaluation of search engines, and subsequent comparisons of results, are complex procedures. If a standard format is not used, results will not be comparable.

## **2.5.6 RELEVANCE, RECALL, PRECISION AND RANKING**

### **2.5.6.1 INTRODUCTION**

Another common thread of comments in the literature covered certain technical aspects of information searching: relevance, recall, precision and ranking.

### 2.5.6.2 RELEVANCE

Relevance can be defined as being the degree to which a retrieved document fits into the user's perception of the best answer possible to his/her original query. Lancaster admitted (long before the first Internet search engine became available) that "... relevance is not a precise term ... " (Lancaster 1978:256). Ingwersen refers to relevance as being "... the Dark Matter problem in IR" (Ingwersen 1992:54). Tonta quotes Robertson, who claimed that "... 'relevance' is defined as a relationship '... between a document and a person in search of information' " (Tonta 2001).

An example from everyday life might clear up the possible lack of clarity resulting from the multitude of definitions. A user asks the question; "Who was the overall winner of the 2000 Tour de France cycling race?", does a search and receives a document which states that Lance Armstrong was the winner. The user will attach a higher value of relevance to this document than to another document which listed only all the individual stage winners for that same event.

The aim of any information retrieval system is to extract relevant information, therefore relevance is a crucial concept in information retrieval (Mizzaro 1998:303, Su 1993:93). Notess stated that initial web search engines made heavy use of relevance to sort output (Notess 1999c:84). However, relevance cannot be quantified – Burgin quoted other authors who have proved the existence of a sizeable difference in agreement on the relevance of a returned set of documents between the authors of the original queries and non-authors (Burgin 1992:619). In fact, Robertson found as far back as 1977 that the variations in relevance judgements for retrieval tests "...are difficult to assess" (Robertson 1977:251). Furthermore, other research proved that an inverse relation exists between subject knowledge and quantity of relevance judgements – more knowledgeable judges will consider fewer documents to be relevant (Burgin 1992:621). Mizzaro described a four-step series of events originating from a user's real information need, called the RIN. The user then converted it to what he/she perceived the need to be, called the perceived information need (PIN).

This PIN was then expressed as a request for information, normally in the user's home language. The request was then formalized, sometimes with the help of an intermediary, into a query using the language of the retrieval system being used (Mizzaro 1998:306). A sweeping claim was made that search engines are easy to use since their operation is based on relevancy searching (Courtois and Berry 1999:39).

A number of other authors' views of relevance are summarized below.

Barry did a study on user-based relevance criteria, which indicated that users employ tangible characteristics (information content, reference provision), subjective qualities (agreement with information) as well as situational factors (time constraints) to make relevance decisions (Barry 1994:149).

Park identified a need that user-based relevance be developed to the user's benefit, and also to allow future research into IR (Park 1994:135). Su found in a study with 40 end-users that there is a high concern amongst users for a high degree of recall, and that a high degree of precision does not always map to high quality answers (Su 1994:207).

Although a large number of studies have been done on relevance in IR, Harter claimed that the effect of variations in relevance assessments on the measurement of retrieval effectiveness is almost unstudied. It was suggested that a series of rigorous tests be done to determine how these variations affect these measurements (Harter 1996:37).

In probably the most extensive study on relevance in IR to date, Mizzaro inspected 157 papers on this topic spanning the period from the 1950s up to 1997. Four kinds of papers were identified: Conceptual, Experimental, Survey and Theoretical. A very basic truth about relevance was restated, namely that relevance is subjective – different judges could have different opinions on the relevance of a certain issue (Mizzaro 1997:810,812,815).

Another study was done by Barry into the role document presentation plays in the relevance judgement process. In conclusion it was mentioned that document presentations differ in their effectiveness as pointer to possible relevance since different types of document presentations do not present clues to their own relevance the same way (Barry 1998:1293).

Saracevic identified five distinct types of relevance in a general study on information science:

- system or algorithmic relevance,
- topical or subject relevance,
- cognitive relevance (pertinence),
- situational relevance (utility), and
- motivational (affective) relevance (Saracevic 1999:1059).

Hirsh referred to a number of factors, as well as the IR process itself, which have an effect on the relevance judgement. In this study, it was found that children change their relevance criteria as the search for specific information progresses (Hirsh 1999:1265,1280).

Spink and Greisdorf found that more items were rejected than accepted in terms of their relevance, based on their topicality (Spink and Greisdorf 2001:161). In general, relevance is one of the most important issues in the study of IR, but one which eludes clear definition and understanding more than most other concepts.

#### 2.5.6.3 RECALL

As with relevance, a number of definitions exist for recall in IR. Recall is a measurement of "... the system's ability in retrieving all the relevant documents in response to the user's queries, questions or problems" (Su 1993:95).

Lancaster supplies a formula for the calculation of the recall ratio as being the number of relevant documents retrieved, multiplied by 100 and divided by the total number of relevant documents in the collection (Lancaster 1978:112).

Ellis claims that recall can be calculated "... by dividing the number of documents in the collection held to be relevant to a question by the number of those documents actually retrieved ..." (Ellis 1996:7). Tonta states that "... recall is defined as the proportion of relevant documents retrieved ..." (Tonta 2001), while other authors define it as follows: "Recall is the fraction of the relevant documents ... which has been retrieved ..." (Baeza-Yates and Ribeiro-Neto 1999:75).

An example is in order to clarify the matter:

A collection of 1000 documents contains 100 on ecology. A retrieval system returns 80 documents, of which 60 are on ecology. The recall of the system is  $60/100 = 60\%$  (Large, Tedd and Hartley 1999:163). The ideal recall figure is 100%, which is achieved when the system retrieves all 100 documents on ecology.

Recall cannot be calculated exactly in real life, since it is impossible to determine the exact number of documents available on the Internet on a given topic, viz. 100 in the example above (Clarke and Willett 1997:185). The other figure, the number of relevant documents (60 in the example above) is easier to determine - once again subject to user judgement. Recall is a relevance-based measure, since it affects the final measure of relevance (Su 1993:93).

Buckland and Gey did a study on the perceived inverse ratio between recall and precision in IR. It was found that the following method could be used effectively on large databases, to increase recall and precision simultaneously; first do an initial search while emphasizing high recall. Secondly, filter the retrieved set of answers. However, the intrinsic tradeoff between recall and precision still remained (Buckland and Gey 1994:12).

In another study, Su proved that there is no significant correlation between the precision of retrieved data and the user's judgement of success (Su 1994:207).

#### 2.5.6.4 PRECISION

Once again a number of definitions exist for the precision concept in IR.

Lancaster supplies a formula for the calculation of the precision ratio as being the number of relevant documents retrieved multiplied by 100 and divided by the total number of documents retrieved (Lancaster 1978:112). Precision can be calculated "... by dividing the number of relevant documents retrieved by the number of documents retrieved ...", claims Ellis (Ellis 1996:7). Tonta states that "... precision is defined as the proportion of retrieved documents which are relevant ..." (Tonta 2001), while two other authors claim that "Precision is the fraction of the retrieved documents ... which is relevant ..." (Baeza-Yates and Ribeiro-Neto 1999:75). Finally, Su defines precision as "...the proportion of relevant references contained in the search results..." (Su 1993:95).

Example:

A collection of 1000 documents contains 100 on ecology. A retrieval system returns 80 documents, of which 60 are on ecology. The precision of the system is  $60/80 = 75\%$  (Large, Tedd and Hartley 1999:163). The ideal precision figure is 100%, which is achieved when all the documents returned by the system are on ecology.

Precision is relatively easy to calculate, since the two figures required (80 and 60 in the example above) are easily obtainable (Clarke and Willett 1997:185).

Feldman suggests that a successful searcher should build up a query to achieve precision without omitting relevant results (Feldman 1999:63).

Precision and recall cannot be separated when relevance is considered, since both affect the final measure of relevance (Su 1993:93).

Leighton and Srivastava evaluated five major search engines by comparing the precision among the first 20 results returned for each of 15 different queries.

Results proved that AltaVista, Excite and Infoseek produced the highest degree of precision, while Lycos and Hotbot performed better with short/unstructured and structured queries respectively (Leighton and Srivastava: 1999:870).

#### 2.5.6.5 RANKING

Ranking refers to the way in which a retrieval system lists the answers it produces as a result of a user query – the most relevant one(s) should be listed first.

Example:

A user states his information need as being “I need beginner tutorials on HTML version 4.0 webpage programming”, and then does a search and receives *p* documents as answer. If the retrieval system has an effective ranking system, it should first list the *q* documents being HTML version 4.0 webpage programming beginner tutorials. The next *r* documents could be advanced HTML tutorials, the next *s* documents examples of HTML code, the next *t* documents tutorials on Java programming, etc.

Since most users only view the first 10 or 20 results of a search, ranking is of extreme importance (Courtois and Berry 1999:43, Notess 1999c:84). The following factors are used in search engine algorithms to determine ranking: number of matched terms, term proximity, term frequency, document length and other factors (Courtois and Berry 1999:39-41, Notess 1999b:64). Feldman adds weighting and term rarity to the list (Feldman 1999:67).

Unscrupulous authors of webpages wanting to increase the ranking of their webpages, use spamming techniques, which include repeating key words in the HTML source coding meta-tag section of the header. Brandt calls this process “loading” (Brandt 1996:39). They could also code many other pages (called jump pages) containing links to the real page, in the hope of lifting the ranking of the real page (Chun 1999:140).

Repeating descriptive words in the smallest size (making them invisible to the reader, but not to the visiting crawler), and using the same foreground and background colours, are other known spamming techniques (Notess 1999c:85).

### 2.5.7 MULTIPLE INDICES

A traditional searcher could refer to more than one different source (books, journals, newspapers, encyclopaedias) for information. A searcher using any electronic medium has the same option. He/she could do a number of sequential searches for the same information using a different search engine or database in each case. This would yield a mixture of old and new answers after each search.

Spink has confirmed that some users (without any advice or training) are doing multiple searches on OPACs and CD-ROM databases for better search results: 57% of a sample of 200 academic end-users conducted multiple search sessions (Spink 1996:603,606).

The value of this richness of information is confirmed by Chun quoting Reid in saying that the more pieces of information available to choose from, the higher the chance of finding the correct one (Chun 1999:137). The fact that those pieces of information could be spread over more than one database, assuming that the databases are all of the same quality, does not affect this basic truth.

However, a breed of search engine called a meta search engine makes this multiple searching possible in one Internet searching session. A meta search engine simply takes the user's query, and submits it to the index of a number of search engines simultaneously. Once the results have been received, website summaries are combined and one set of answers is produced.

Garman in fact built a case for the use of meta search engines by strongly advising users to search multiple databases to satisfy the same query (Garman 1999:75). Other authors confirmed the validity of this advice (Jacso 1999a:227, Jacso 1999b:99, Mickey 1999:79, Williams 1996:246).

Baeza-Yates and Ribeiro-Neto describe parallel and distributed IR as an application of hardware-based parallel computing, based on a well-known variety of processor configurations: SISD, SIMD, MISD and MIMD (single instruction stream single data stream, single instruction stream multiple data stream, multiple instruction stream single data stream and multiple instruction stream multiple data stream) (Baeza-Yates and Ribeiro-Neto 1999:230). A parallel is evident between this pure hardware implementation of multiple searches and the software implementation as used by meta-search engines.

### 2.5.8 SEARCHING SUCCESS RATE

Information plays a critical role in any organisation (Evans and Wurster 1997:72). Initially, traditional information retrieval was considered as a topic separate from the Internet: "This (the Internet) is not about information retrieval as we know it. ... But it's really not about information retrieval as we have historically tended to think about it" (Dunn 1996:77).

Lancaster discussed some factors which influence success in online searching many years before the advent of Internet search engines (Lancaster 1978:176). However, lately many authors agree that learners should know how to retrieve information from electronic sources (Cronje and Clarke 1999:214, De Jager and Sayed 1998:197, Edling 2000:11, Gan 1998:29, Marais and Marais 1999:85-86, Wong 1998:5). IS managers consider the process of letting users (i.e. what the learners of today will become) find information for themselves as liberating to IS staff (Foley 1996:45).

Some authors claim that, in general, finding relevant information in an information source is not a difficult task:

- "... the majority of Australian academics have a high expectation of success as they engage in information seeking on the Internet" (Bruce 1998:553).
- "... Australian academic users generally ... are satisfied with information seeking..." (Bruce 1998:555).

- "... most children could successfully seek and retrieve information in response to information related problems using a range of information media." (Oliver and Oliver 1997:529).
- "... even the novice information searcher is able to extract relevant documents in rank order of calculated relevance." (Ross and Wolfram 2000:949).

However, a large number of authors claim that the opposite is true:

- "Finding relevant and precise information on the web can be extremely difficult" (Coetzee 1999:9).
- "Unless students think critically about the tools they use to locate resources, the results may well be disappointing, if not downright misleading" (Cohen and Jacobson 1999:1).
- "The rich promise of Internet information is tantalizing, but usually frustrating" (Huwe 1999:68).
- "The very size of digital libraries begs the question of how average users will continue to find context and meaning." (Huwe 1999:68).
- "Would a savvy searcher know all these (a large number of semantic and syntactic variations for the same concept)? Hardly. A casual user such as a college student? No way" (Jacso 1999b:99).
- "Increasingly user-friendly interfaces are making it more difficult to display searching principles" (Keenan 1997:237).
- "Using Internet search engines to retrieve information can sometimes be a slow process ... and can result in far too many 'hits' so that picking out relevant ones can be difficult" (Large, Tedd and Hartley 1999:5).
- "Our customers (where customers are library users) understand their information needs, but they don't understand the tools that can satisfy them" (O'Leary 2000:22).
- "Currently, search is simply bad" Sherman (1999:54) quoting Truher.
- "...some respondents seemed confused about what they were to report when asked to list query terms for their search" (Spink, Bateman and Jansen 1999:122).

- "...the user's ability to specify good search terms and create complex search queries to clearly and precisely capture relevant retrieval seems rather low." (Spink, Bateman and Jansen 1999:125).
- "As the World Wide Web grows exponentially, discovery and retrieval of useful educational material grows more problematic" (Sutton 1999:1191).
- "Only 33% of the Internet users agree or strongly agree with the statement 'It is easy to perform subject searches on the Internet.'" (Voorbij 1999:604).
- "I find it difficult to search information on the Internet..." (Voorbij 1999:605).
- "...information seeking is a complex and difficult process for these students, who seek to reduce the task to finding an obvious answer or finding a good website..." (Wallace, Kupperman and Krajcik 2000:75).
- "... both novice and experienced searchers were overconfident in their performance" (Wolfram and Dimitroff 1997:1145).

Furthermore, Jones criticizes web robots as using crude methods of gathering data, not distinguishing between opinion and fact (Jones 1996:80).

Both the positive and the negative set of comments above were extracted from the same set of references. This would seem to imply that a comparison of numbers should indicate the weight each one carries: four out of 20 versus 16 out of 20. Thus 20% of these authors consider searching the Internet relatively easy, while 80% consider it to be relatively difficult. After having been involved with consultation, training and research on Internet searching for a number of years, the author strongly agrees with the larger group. It has been his experience that most average computer users experience intense difficulty in finding relevant information quickly on the Internet. From personal experience (without having kept exact records), the author can confirm many of the statements made by some authors quoted in this research, including:

- searchers have trouble in formulating search queries;
- searchers often use a single word as search query, and often receive thousands and even millions of apparently irrelevant answers;

- searchers seldom make use of Boolean and other more advanced filtering functions, and;
- searchers are overwhelmed by terms such as relevance, query, ranking, URL, operator, proximity, etc.

In a study done on the information-seeking patterns of artists and art scholars, Van Zijl and Gericke found that the Internet was least popular on a list of six possible information sources. Searching the Internet as a method of information retrieval was also last on a list of three choices (Van Zijl and Gericke 1998:23).

Finally, Tate and Alexander state that the quality of information on the web is suspect owing to the absence of editors and fact checkers (Tate and Alexander 1996:49).

The sheer number of authors claiming failure or at least difficulty during a process of Internet information searching seems to indicate that finding relevant information on the Internet is not an easy task. This fact strengthens the case for this research project.

## **2.6 SEARCHING METHODOLOGIES**

As with any process, information seeking requires a structured methodology to ensure success. Even early authors elaborated on searching strategies to be used in IR long before the advent of the Internet (Lancaster 1978:156).

However, since there are so many variables involved, this methodology cannot simply be described as a series of steps which will guarantee success.

Marchionini did however define a sequence of steps followed in IR, with a number of complex relations between them: Problem definition, Source selection, Query formulation, Query execution, Result examination, Info extraction and Reflection (Marchionini 1995:50).

Other authors attempted to verbalise this process as follows: "Information seeking can be viewed as either information searching or information browsing ... although information seeking is often, in practice, a combination of these two activities" (Large, Tedd and Hartley 1999:143). A large number of authors have done studies on the intricacies of searching methodologies.

One of the earlier studies done on users searching for electronically stored information, starts off by stating a belief that end-users will one day do their own searching, provided search processes have been simplified (Fidel 1986:37). As a result of the openness of the Internet, the computing community has certainly reached that point. A detailed map is given, with successive actions to be taken in an attempt to search successfully, called the Selection Routine (Fidel 1986:39). This author was convinced that this routine would enable searchers to select the most appropriate search key, leading to success in searching (Fidel 1986:42).

The same author did an in-depth study in 1991 on the selection of search keys by searchers, described in three articles. In the first, a case study was done with 47 professional online searchers, proving that a formal model of searching behaviour can be created even with a relatively small number of searchers (Fidel 1991a:499). In the second article, results proved that searchers based their key selection on thesauri and indexing, and that multi-database searches prescribed the existence of high quality thesauri and the switching from one language to another (Fidel 1991b:501). In the third article, an important distinction between the way searchers altered their methodology while searching came to light: so-called operational as opposed to conceptual moves (where a move is a modification of a search strategy) (Fidel 1991c:517). An operationalist searcher attempts to apply the most optimal search strategies to achieve precise retrieval. During their interaction, they maximise the system capabilities but retain the specific meaning of the original search request. Conceptualist searchers however, fit a specific request into a faceted structure, and are mainly concerned with recall. They may also change the specific meaning of the original search request in an attempt to find the best results (Fidel 1982).

Conceptualist searching has a parallel in some of the work done by Lancaster, who describes conceptual analysis as those aspects of a document which are of most concern (Lancaster 1978:9).

Large and Beheshti quotes Borgman who lists three layers of knowledge required by a searcher:

- conceptual (convert information into a query),
- semantic (construct a query) and
- technical (enter query as specific statements) (Large and Beheshti 2000:1079).

There is a clear similarity between these layers of knowledge and the operationalist versus conceptualist approach of Fidel, as discussed above.

According to a study by Kuhlthau, a searcher passes through six stages during information searching, with certain emotions and thoughts characterizing each phase: initiation, selection, exploration, formulation, collection and presentation (Kuhlthau 1991:366-368). This author also reasons that the very nature of information seeking is such that potential anxiety could be present (Kuhlthau 1991:370). It was found that there is a difference between the way a user approaches information seeking, and the way information is provided by the systems offering these information seeking services (Kuhlthau 1991:361).

Fidel defined document-oriented indexing as a method whereby documents are indexed based on their contents, while user-centered indexing should indicate the requests which might be relevant for that document (Fidel 1994:572).

However, it was found that research on user searching behaviour was still too immature to develop user-centered indexing (Fidel 1994:576).

A study was done with experienced searchers using online catalogues for subject searches. Four kinds of knowledge were defined: factual, experiential, process and meta-knowledge.

The last type was defined as being the integration of the other three, but also being a volatile, ever-changing kind of knowledge (Connell 1995:509). It was noted in conclusion that the environment does have a significant effect on the searcher and thus the search results (Connell 1995:517).

Green defined relevance as the relation between a piece of text and the information need a user might have (Green 1995:652). However, this study found that further research would be required to determine what subset of relationship types actually prescribe topical relevance (Green 1995:646).

Another study by the same author attempts to provide an answer to this question. It was found that the topical relevance relationship, similar to the concept of relevance, is a highly complex concept, and requires further research (Green and Bean 1995:662).

In a 1996 study, Harter and Cheng reported on the concept of using co-linked descriptors, which are alternative search terms to those originally chosen by the searcher. This approach appeared to be effective, although methodological problems plagued their implementation (Harter and Cheng 1996:311). Oliver and Oliver proved that high levels of learning and knowledge acquisition take place during searching activities where the context and purpose of the activity is known. The perception that learning automatically takes place during the execution of an information seeking task was proven to be false (Oliver and Oliver 1997:519,525).

Very large databases are being introduced into the computing environment, where the relation between various sets of data is not explicit in the database structure (Vickery 1997:107). This author introduced new procedures to extract knowledge from such databases, and gave an indication of how these methods could be used (Vickery 1997:119).

An empirical study was done by Yuan to investigate the effect of the experience level of an end-user searcher on his/her searching behaviour. The searching behaviour of a group of law students was monitored over a one year period.

It was found that search experience did have an effect on the following aspects of their behaviour: increase of command and feature repertoires, language usage pattern, increase of search speeds and change of learning approaches. However, the amount of experience had no positive effect on the number of errors being made or on the recovery from those errors (Yuan 1997:218, 230,231).

Chen and others did an empirical study to test two algorithms which could help the user in his/her searching approach to information retrieval. These algorithms were a category map to assist in browsing, and a concept space searching algorithm. It was found that both methods had a substantial effect on the success factor of searchers (Chen and others 1998:582,600,601).

Hirsh found that elementary school children used sophisticated search techniques during an empirical study, but showed little concern for the validity of references found on the WWW (Hirsh 1999:1265,1270).

Large, Tedd and Hartley identify two basic approaches to the ideal situation of finding the exact and correct result to an online search: search-narrowing tactics when a search produces too many possible answers, and search-broadening tactics when a search produces no or few, unrelated answers (Large, Tedd and Hartley 1999:166,168).

Large and Beheshti studied 50 grade six primary school children while they were using the WWW as information source for classroom projects. As a result, the children were divided into one of three groups, based on their behaviour while searching:

- technophiles (expect the best possible results from the WWW, even though their experience with actual results does not necessarily endorse this expectation),
- traditionalists (prefer printed references such as encyclopaedias), and
- pragmatists (use both WWW and traditional sources, depending on the kind of information sought) (Large and Beheshti 2000:1072).

It was clear that the methodology used by searchers is not easily definable, does not always produce acceptable results and is not easy to measure. Fidel's work seems to suggest that a searcher should spend energy on the conceptualist rather than the operationalist approach. If a model is to be designed which attempts to assist learner searchers, its focus should thus be away from operators and other technical detail. It should rather lead the user to think more about the topic itself, and not concentrate on the detailed methods and operators involved in the actual search itself. This author decided to build the final searching model on this basis, allowing the user to concentrate more on his/her topic than on operators, URLs and other technical details.

## **2.7**            **SUMMARY**

It appeared that traditional methods used by educators may have to be altered to incorporate new technologies, and to meet the expectations of learners. Unfortunately, research in this area is still relatively immature.

One point that was established beyond doubt by numerous authors, is that finding a specific piece of information on the Internet is not an easy task. One possible reason is the vastness of the mass of unorganized information available on the Internet. In parallel with this fact it also became clear that more than one index should be used to ensure more accurate answers from searches.

There are many tools in the form of search engines, each one with a set of operators, available to assist users in finding information. Many evaluations and tests on these tools have been done. An obvious question arose at this point: if there are so many searching tools with operators available, why is it difficult to find relevant information on the Internet? This and other questions will be addressed in this research project.

# **CHAPTER 3**

## **LITERATURE ANALYSIS**

### **3.1 INTRODUCTION**

The purpose of this chapter is to analyse the literature discussed in Chapter 2. The same numbering scheme will be used, to make cross-referencing easier. For example, the literature reviewed in Section 2.2.1 is analysed in Section 3.2.1.

### **3.2 HISTORY AND STATUS QUO**

#### **3.2.1 INFORMATION SCIENCE AND INFORMATION RETRIEVAL**

It is clear from the literature that IR has a rich history, which spans five and a half decades. It does, however, contain some difficult to define, contentious components, of which relevance is possibly the most outstanding one. It was interesting to note that some descriptions of concepts as early as 1945 have a strong parallel in modern WWW techniques. Bush's suggested interaction between user and information store serves as an example (Bush 1945).

Most authors agree that IR has become an important role player in the information industry. The correlation between a point noted by Saracevic and the second part of the statement by Calvin Mooers (see Section 2.2.1.2) is clear: IR cannot be discussed without considering the role of the technology used to make it possible (Gupta and Jain 1997:71, Saracevic 1999:1058, Spink and Losee 1996:43).

However, technology is not the only factor to be considered in IR. To date, Mizarro's in-depth study on this all-important concept of relevance stands out as the most comprehensive collection of research work.

However, the claims that relevance is better understood than ever before, but that the definition of the term is a challenge, points to a basic contradiction (Schamber 1994:36). Since relevance is an intricate and subjective measure, it is doubtful whether a concise definition will be found which will satisfy most IR experts. Greisdorf quoted Goffman, who attempted to set up such a definition: (Relevance is) "... a measure of information conveyed by a document relative to a query" (Greisdorf 2001).

The Cranfield tests, although dated, conclusively proved that simplicity in IR is a solid design principle. They also suggested that, although the tests were essentially of manual systems, their focus on the importance of retrieval characteristics as perceived by the user should remain an important guiding principle in the design of automated information retrieval systems (Ellis 1996:2).

While the information explosion triggered the growth of the WWW, IR has grown in parallel with the technologies carrying the WWW. Where the WWW provides the channel to information, IR provides the searching mechanism to it. IR has reached the point where it could lead to strategies for dealing with parts of the problem of retrieving relevant information from the Internet. The aim of this research project is to develop one such strategy.

### **3.2.2 THE INTERNET AND HYPERLINKS**

Two clear development processes in the building of the Internet are evident from the literature. On the one hand, a solid base of hardware and protocol standardization has been woven into the internal structure of the Internet. The establishment and use of the TCP/IP network protocol as basis is an example. This resulted in a measure of consistency, for example in the way browser programs can access websites, regardless of which operating system platform both the browser and the website are using. In the second place (and on a less positive note), a total lack of control and management of Internet information content stands out as a very prominent feature. Some of the events leading to this situation include the creation of e-mail and the Usenet feature (Kristula 2001).

Both have different objectives and characteristics, but both added to the massive volume of data generated by the users and are stored somewhere on an Internet server. However, searching for relevant items in both these forms of communication is also an important part of an IR system. Unsolicited e-mail and its effects on productivity in itself are large enough problems to warrant further research. A literature survey and empirical study by Morrison shed some light on this issue (Morrison 2000).

The existence of the useful and successful concept of hyperlinks also assists in the uncontrolled access to unwanted information (Bush 1945, Butner 2001). A user could easily link from one website to another, possibly ending up with unwanted materials. Although standard printed materials do offer similar freedom to a potential reader (simply by being more accessible in a global sense), the effortlessness of "moving" between documents offering a wide variety of information offered by the WWW is unparalleled.

Initially, Gopher and similar applications provided a supposedly familiar way of navigating electronic resources. However, the first and subsequent versions of browser programs actually attempted to bring a degree of order to the user part of the chaos of the Internet information world (Gribble 2001). At least the user could now navigate the immense Internet landscape with more ease than before. However, it made no difference to the fact that the actual amount of information was growing apace, and that there was still no control over its structure, content or access.

The exponential growth of the Internet poses a variety of problems to various disciplines. For example, user complaints over access speed never seem to decrease, regardless of new capacity for Internet traffic being added continuously. The author has experienced this very fact intensely at his home institution. Furthermore, the growth of the user base creates problems for Internet service providers to keep up with providing unlimited access via dial-up telephone lines (Kristula 2001).

It is clear from the literature that the large amount of information available, and the fact that it is not centrally indexed, pose serious problems to the IR field (Chun 1999:140, Sherman 1999:54,57). It is debatable whether it is possible or even desirable to attempt to index the Internet in one central location. If it is done, the question arises whether the index should consist of assigned or derived indexing, or both. By default, search engines create indices based upon the complete, or part of, a document, thus the majority of terms are derived. The Cranfield experiments of manual indexing systems concluded that the inclusion of assigned index terms made no statistically significant difference to retrieval performance but did add significantly to costs. The economic argument is one of the high cost (human time) but high quality of assigned indexing versus automated, cheaper derived indexing done by machine. Owing to the sheer size and growing rate it would appear to be an impossible task to index the whole Internet.

One potential solution appears to be the use of standard search engines, although they only index a small section of the information available on the Internet. Lawrence and Giles found (in a study published in 1999) that no search engine covers more than 16% of the estimated size of the publicly indexable WWW. It was also found that only 34% of websites make use of the most basic meta-tags to enhance indexing ("keywords" and "description"), while only 0.3% used the Dublin Core meta-standard (Lawrence and Giles 1999).

It was concluded that search engines should play a part in the proposed solution to the problem of access to academic materials via the Internet.

### **3.2.3 IT IN EDUCATION**

It was clear from the literature that the traditional roles of the educator and the learner were changing. Traditional teaching methods also face an onslaught of modern technologies, mostly made possible by IT.

Some authors suggested that the emphasis in education should shift from the educator to the learner (Ives and Jarvenpaa 1996:35).

Others believe that educators cannot afford to retain the same methods and media to convey concepts to learners as they have been using for many years. Specifically the role of the following (as the one and only method of instruction) should be reconsidered: face-to-face instruction; an environment where the learner remains passive; one-way communication from educator to learner; and parrot-fashion regurgitation of knowledge (Ives and Jarvenpaa 1996:35, Davis 1996:849,852, Marais and Marais 1999:88). All four of these options emphasize the role of the educator while de-emphasizing the role of the learner, and they reduce the opportunity for the learner to exercise his/her creativity - as a result less learning could take place.

Alternative methods and media of instruction include the use of computers, the Internet, communications, computer laboratory practicals and distance education using IT. One author proposes in strong terminology that, if used, these methods could revolutionise education:

- "Telecommunications in general and the Internet in particular are potentially the most significant educational tools that have become available ..." (Maddux quoted by Selwyn 1999:225).

Other authors suggest that these methods are already in use:

- "... technology-based teaching methods ... including the Internet, e-mail, cable and satellite broadcasting of lectures, and many more will have to be introduced" (Marais and Marais 1999:88).
- "... utilisation of the knowledge media on the Internet has begun to revolutionise higher education" (Cookson 2000:72).

These views have an implication on quality control in education in general, and specifically in higher education. Some authors are questioning the effectiveness of the transmission of knowledge and skills, which is supposed to take place. This is not the beginning of the debate – the effectiveness and the objectives of teaching have been questioned from the nineteenth century onwards.

A new angle that comes to light now is the way in which the use of technology may assist or influence the effectiveness of this transmission of knowledge. A question that comes to mind is: where does the responsibility for the effectiveness of the learning process lie?

It is the view of the author that the answer is three-fold: the educational system should empower the educator to involve the learner in his/her own studies in such a way that learning will take place in a more natural fashion than with rote studying of knowledge only. This view stems from many years of working inside an education system which, as result of a long tradition, prescribes by virtue of the way it operates and its resultant momentum that both educator and learner should act in a certain way. The author believes that an education system can produce better graduates if its participants have mastered the skill of learning as opposed to the skill of studying. Finding up-to-date academic resources on the Internet is considered to be part of the skill of learning.

It is thus concluded that educators wishing to survive in the educational world will have to view new technologies critically, and strongly consider adapting to the relevant ones in their teaching.

Contrasting views exist regarding the success of the application of IT in education. Some authors report only positive results after having applied IT in some form to their educational environments. Some of the advantages listed include higher quality of education, lower cost, more work covered in the same time, deeper coverage of work and expanding the number of courses offered (Cookson 2000:73, Edling 2000:10,13). Others are more cautious and warn against using IT blindly to further educational goals, since research in the classroom use of IT is immature (Wallace, Kupperman and Krajcik 2000:75). It was also noted that past predictions of technology replacing books have been proved false, and that IT should be applied with care in education (Gan 1998:32, Mellon 1999:28,29).

Another group reported negative results. Research into IT in education is still young, and a tendency exists to apply IT to any problem situation in education, hoping it will cure it, regardless. Some reports claimed actual failure of the use of IT in the school environment, possibly due to money being spent on hardware and software for economical rather than educational reasons. Other reasons listed include a shortage of information professionals and a low level of basic computer literacy (Saracevic, Braga and Afolayan 1984:192, Selwyn 1999:227).

The conclusion drawn here was that educators and management alike should approach any new direction in education, where IT is to be used for the first time or in a new way, with care. Research results could play a vital role in guiding management decisions. IT should not be "applied" to a problematic educational situation without planning, in the hope that it will magically solve all problems. The author took note of these factors in an attempt to ensure success in the design, testing and application phase of the proposed searching model.

### **3.3 THE LIBRARY**

On the one hand Voorbij claimed that libraries should support their users by assisting them to find information, while Basu, and Devlin and Burke found that librarians had trouble in supplying information to users from electronic sources (Basu 1995:38,39, Devlin and Burke 1997:104, Voorbij 1999:598). The solution to this apparent contradiction is provided by Saracevic, who claims that librarianship and Information Science are related but different fields (Saracevic 1999:1060).

#### **3.3.1 THE ROLE OF THE WWW**

Since both libraries and the Internet store information and deliver it on request to users, it is a natural consequence that the roles of these two entities are weighed against each other. Lewis described the apparent danger posed to libraries in the 1980s, when the rising cost of traditional publishing was contrasted with electronic publishing (Lewis 1980:63).

Rather than viewing the two as being in opposition, libraries can be seen as facilitators to access to information while the Internet is an additional, complementary channel of information supply.

It is clear that in many instances libraries have adapted on a large scale to integrate their services and the Internet in some way (Furner-Hines and Willett 1995:31). In other cases, however, librarians struggle to meet the needs of information seekers (Basu 1995:38,39, Devlin and Burke 1997:104).

The conclusion reached is that libraries could provide an even better service to their users by integrating their existing information provision methodologies with the information provision services of the Internet. This should serve to get even closer to the ultimate goal of satisfying the users' information demands.

### **3.3.2 THE WWW AS SOURCE OF EDUCATIONAL INFORMATION**

There is no doubt that the Internet has a wide range of sources of educational information, but that it cannot be assumed that learners can easily locate them (Bar-Ilan 2000:432, Fidel and others 1999, Hirsh 1999, Large and Beheshti 2000:1069, Oliver and Oliver 1997, Spink and Saracevic 1997, Sutcliffe, Ennis and Watkinson 2000, Voorbij 1999:604). Furthermore, it has been proved that it also cannot be assumed that learning takes place during the location process (Oliver and Oliver 1997:519). In fact, failure in identifying answers and resultant frustration appear to be commonplace among these searchers (Large and Beheshti 2000:1079, Sutton 1999:1182).

The relevance issue needs to be mentioned at this point. If a searcher claims that relevant information is found or not found, on what basis was the relevance judgement made? To address this problem, the author decided to define an area of knowledge within which each participant in the searching experiments to follow would feel at ease. This would be achieved by asking each searcher to choose an information need, which appeared as a direct result of their current academic studies.

Since the searching experiments would be done close to the end of the academic year, it was assumed that learners would at least be familiar with the subjects they were studying, and would be in a position to judge whether or not the answers they found fulfilled their information needs.

These points strengthen the motivation for this research project. If educational information on the Internet exists in large quantities, but is difficult to locate, a way should be found to assist learners in this quest.

### **3.4 CASE STUDIES AND EXPERIMENTS**

#### **3.4.1 INFORMATION RETRIEVAL**

From a number of empirical studies involving actual experiments, it was clear that successful IR depends amongst other things, on a de-emphasis of the role of technology in the searching process. This became clear in the study on the sixth-graders who focussed on the computers rather than on the searching process (Wallace, Kupperman and Krajcik 2000:75). A number of other relevant points also surfaced:

- It appeared that searchers did not view the searching process as a learning experience.
- A short training session minimised the negative effect of searching technicalities.
- Original search terms produced highly relevant answers.
- Many searchers experienced problems in retrieving subject-specific material from the Internet.

It appeared that a short time spent on training before a measurement of success in searching, could yield a large difference in the level of success (Hersh, Pentecost and Hickam 1996). These points were considered to be highly relevant to the current study, and would be incorporated in the design of the proposed model, especially the value of a short training session.

### 3.4.2 SEARCHER BEHAVIOUR

A number of case studies were inspected. This included some non-Internet searching studies, which proved that:

- the use of search tactics was affected by search experience (Hsieh-Yee 1993:161, Yuan 1997:231), and
- Boolean operators were seldom used (Siegfried, Bates and Wilde 1993:273).

The first point above confirms the basic theory of Expectancy, as defined by Vroom: when an individual has to make a choice, his/her behaviour is affected by the degree to which he/she believes these outcomes to be probable (Vroom 1964:17).

Other studies involving Internet searching proved that:

- searchers were generally satisfied with their results, but complained about slow responses (Fidel and others 1999:24),
- in general, searchers had problems in formulating proper search queries (Hirsh 1999:1268), and
- some sophistication in searching techniques was evident, after participants were given some basic training (Large and Beheshti 2000:1076).

The finding about participants being satisfied with their results above is contradictory to a large number of other studies (Coetzee 1999:9, Huwe 1999:68, Large, Tedd and Hartley 1999:5, Spink, Bateman and Jansen 1999:122,125, Sutton 1999:1191, Voorbij 1999:604,605, Wallace, Kupperman and Krajcik 2000:75, Wolfram and Dimitroff 1997:1145). However, the sample of searchers in the 1999 Fidel study consisted of only seven high school students. This contradiction is ascribed to the small sample size, since participants could have been indiscriminating or unaware of how to assess information.

Based on the findings in Section 3.4.1 and 3.4.2, a number of decisions were taken with regard to the proposed model:

- A short, basic training session would be done on the use of the model.
- The searcher would be requested to specify his/her original search query fully.
- The sample size of participants in searching experiments would be based on calculations suggested by prominent authors, to ensure that experimental errors were minimized.

### **3.5            SEARCH ENGINES**

#### **3.5.1            INTRODUCTION**

The use of a search engine is central in this research project, and a number of research results from other researchers' work in this area proved to be useful. The proposed model aimed to make the finding of Internet-based information easier. As a result, it was decided to identify some of the features in existing search engine programs, which were proven to be counter-productive, and to ensure that the proposed model circumvented these problems. These features could be related to interface design, choice of operators, query formulation, choice of search engine and use of multiple indices.

#### **3.5.2            INTERFACE DESIGN**

Any computer program should be designed in such a way that the interaction of the user with the program is effortless, intuitive and otherwise conducive to productivity and enjoyment. Unfortunately, search engine authors did not always apply this design principle. Large, for example, considered lack of design as one reason for lack of success in searching (Large and Beheshti 2000:1079). Bilal found the design of the Yahoooligans! children's web search engine interface to be lacking (Bilal 2001:134). It was decided that the final version of the model would be relatively simple in design.

### 3.5.3 OPERATORS

A large variety of search engine services exist, which should make it easy for users to focus their searches and limit their answers to only relevant ones. Each service offers an array of operators to assist users in this focussing process. Common operators include inclusion (+), exclusion (-), phrases ("...") and Boolean operators (AND, OR, NOT). However, since the syntax of search engines differs, it might not be obvious how to employ these operators in a search.

Boolean search operators not only have a long history, but also appear in most search services as a set of available operators to enhance the searching experience (Frants and others 1999:86). However, it was shown that specifically Boolean operators are either not used, or often used incorrectly by information seekers (Siegfried, Bates and Wilde 1993:273). Hirsh found that children experience problems when working with Boolean operators (Hirsh 1999:1266,1271).

Various other problems were found with operators, such as the fact that some functions and operational logic sometimes produced inconsistent answers (Snyder and Rosenbaum 1999:380).

It was concluded that as many as possible of the available search operators would have to be inspected before including any one in the proposed searching model. If at all possible, the user would be exposed to these operators as little as possible.

### 3.5.4 QUERY FORMULATION

Similar results to those under Section 3.5.3 were found in the literature on studies of search query formulation. Many searchers found this procedure problematic, including basic problems such as spelling errors (Spink, Bateman and Jansen 1999:123,125, Su 1993:97, Hirsh 1999:1266, Wallace, Kupperman and Krajcik 2000:87).

However, at least two studies were found which involved the creation of an “automated query generator” system, which apparently freed the user from struggling with syntax rules (Frants and Shapiro 1991:16, Gauch and Smith 1993:124). Although neither study involved Internet searching using search engines, the concept appeared attractive, and would become central in the design of the eventual searching model. As far as possible, the conversion of an information need into a search query would be done automatically.

### **3.5.5 SEARCH ENGINE EVALUATION**

A large number of authors have done some form of evaluation on various elements of Internet search engines. However, no standard evaluation method exists, and it has been impossible to accurately compare the results of these evaluations. General consensus appears to exist that a standard tool for search engine evaluation should be found (Landoni and Bell 2000:124,126,129, Oppenheim and others 2000:190). An attempt was made to find a method, which would produce a better measure of search engine success than established, traditional measures (Losee and Paris 1999:882,888).

The focus of this research project is to design and implement a searching assistant for learners, and not to design a new search engine. This model would not include its own database, nor a method to keep it updated. It would only involve a front end through which the user would submit his/her query to other, existing search engines. However, it would relieve the user from dealing with syntax and complex search operators. As a result, the evaluation of search engines is not directly applicable to this project.

### **3.5.6 RELEVANCE, RECALL, PRECISION AND RANKING**

Relevance is a difficult concept to define, and can be generalized as being the degree to which a retrieved document fits into the user’s perception of the best answer possible to his original query. The “user’s perception of the best answer ... ” is one of the grey areas of this definition, and could yield different answers from different users with identical backgrounds, on the same query. Furthermore, “the degree to which ... ” is another grey area – for example, what is the difference between 80% and 90% relevant?

The Cranfield tests played a large part in preparing the playing field for research and other work on relevance (Cleverdon, Mills and Keen 1966, Robertson 2001, Tonta 2001). Relevance is clearly an affective concept which is impossible to quantify, while the remaining three concepts below are metrics, and they can be calculated.

Recall is a measure of a system's ability to extract only the relevant documents from a collection (Lancaster 1978:112). It is theoretically possible for a system to achieve a 100% figure for recall, since it could retrieve all the relevant documents present in the collection (Large, Tedd and Hartley 1999:163).

Although recall is defined in a way that allows it to be calculated exactly, one of the basic numbers in the calculation is difficult to determine on a closed database, and impossible on an open database (such as the Internet) (Clarke and Willett 1997:185). Approximations have to be made when enough detail does not exist to make more accurate calculations possible. Lancaster's work on the MEDLARS system shed some light on this topic (Lancaster 1978:35,36, Tonta 2001).

Precision is a measure which almost matches recall in its definition, with the difference lying in the fact that the total number of documents retrieved is required (easy to determine), as opposed to the total number of documents in the system for recall (difficult to determine). Since both required figures are easy to determine, precision is easy to calculate (Clarke and Willett 1997:185).

It has been proved that many users only view a small fraction of the results provided by search engines. This number can be as low as only the first 10 or 20 (Courtois and Berry 1999:43, Notess 1999c:84). As a result, search engine authors have to ensure that they rank the output of their programs properly.

Search engines rank webpages according to a variety of factors. Since there are economic implications involved, some webpage authors use unscrupulous methods to improve the ranking of their webpages.

### **3.5.7 MULTIPLE INDICES**

It was suggested that more than one index should be used during Internet searching to ensure reliable results. The value of searching through more than one database for the same information was confirmed by a number of authors (Chun 1999:137, Garman 1999:75, Jacso 1999a:227, Jacso 1999b:99, Mickey 1999:79, Spink 1996:603,606, Williams 1996:246). This could be implemented by using a meta search engine, or by manually searching using more than one standalone search engine for the same query. Both methods in effect have the same results.

It is concluded that the use of multiple indices should play a role in the design of the proposed model, to give the user a wider scope of databases to be queried.

### **3.5.8 SEARCHING SUCCESS RATE**

Some authors noted that finding relevant information on the Internet was an easy task, while a much larger number reported the opposite. At the same time it was noted that searchers generally do not use advanced searching features, which could be partially responsible for the lack of reliable results.

It is concluded that this difficulty in finding information remains the most important driving factor for this research project. However, owing to the contrasting reports on searching success, part of the research will involve determining experimentally what the situation is.

## **3.6 SEARCHING METHODOLOGIES**

A number of studies have attempted to define methodologies which could be applied successfully to increase searching success (Fidel 1986:39, Fidel 1991a:499, Kuhlthau 1991:366-368, Large, Tedd and Hartley 1999:166,168, Marchionini 1995:50). A landmark point in IR was the definition of the operationalist versus the conceptualist approach, as done by Fidel (Fidel 1991c:517).

The former approach stresses the “technology” of the search engine, including searching methods, while the latter even changes the original search specification in an attempt to think around the concept of what the search topic actually is.

Some authors did experiments, with positive results, on an algorithm to assist users in searching (Chen and others 1998:582,600,601). It was concluded that the proposed model should help the searcher concentrate on the topic of the search, not on search engines and syntax.

### **3.7            SUMMARY**

It was clear that although much work has been done on IR in general and from the Internet specifically, many questions are still unanswered. One very basic and prominent point remaining is that many searchers do not find searching easy, and they find it a frustrating and confusing experience to face so many, very different search engines, operators and methods. This research project addresses this issue.

# **CHAPTER 4**

## **RESEARCH METHODOLOGY**

### **4.1 INTRODUCTION**

The literature reading and analysis clearly identified certain trends, problems and situations in IR. Some of the more prominent ones include:

- in terms of number of computers, the Internet has grown at a phenomenal rate,
- as a result, the amount of available information has expanded dramatically,
- the Internet holds promise as an immense store of academic resources,
- many users complain that finding relevant information on the Internet is fraught with difficulty,
- some attempts have been made to make this process easier, and
- search engines do offer a method of mining these resources.

A number of decisions were taken, based on other researchers' studies, on the design of an instrument to assist learners with the finding of information on the Internet. These include decisions about interface layout, training, simplicity and the use of search engines.

#### **4.1.1 OBJECTIVE**

The objective at this stage of the project was to define the hypothesis, population, sample size, control measures and other relevant points.

Furthermore, various research methods had to be inspected, the most suitable one for the task identified, and the research tool had to be designed and tested.

#### 4.1.2 HYPOTHESIS

The following points from prior reading led to the formulation of the hypotheses below:

- 4.1.2.1 The use of Internet technologies in education has many benefits (Cookson 2000:73).
- 4.1.2.2 Young learners found it difficult to use the information seeking facilities of the Internet (Wallace, Kupperman and Krajcik 2000:75).
- 4.1.2.3 Learners were often disappointed in the results of their attempts to search for specific information on the Internet (Cohen and Jacobson 1999:1).
- 4.1.2.4 Emphasis should be put on the acquisition of skills rather than the passing of examinations (Marais and Marais 1999:88).

These authors seem to suggest that although the Internet and particularly its searching facilities could have a positive influence on learning, learners generally have trouble in coping with IR from the Internet.

Behr suggested the null and alternative hypothesis approach, and it is considered applicable here since it is necessary to measure the eventual success or failure of the mechanism to be developed (Behr 1988:59). The following null hypothesis is postulated:

**H<sub>0</sub>: IT/IS learners can use the searching facilities on the Internet effectively without any aid to find reference materials, which will assist them in their studies.**

An alternative hypothesis has also been formulated:

**H<sub>1</sub>: IT/IS learners cannot use the searching facilities on the Internet effectively without any aid to find reference materials, which will assist them in their studies.**

At this stage it was decided to attempt to prove one of the hypothesis by a series of experiments, as broadly outlined by Leedy:

- "Test the subject..."

This would be achieved by allowing participants to do Internet searching in a controlled environment, without any assistance, and by recording their success/failure.

- "... do something to the subject..."

Next, the participants would be exposed to the use of the searching model during a short training session.

- "... test the subject again to see if there is a difference..."

The first searching session would now be repeated, but this time by using the search model. Finally, the success ratio without and with using the model respectively (for the same participant) would be compared. If a significant increase in the success rate were found, the null hypothesis would be considered false, and the alternative hypothesis would be accepted as true (Leedy 1985:210).

#### 4.1.3 MODEL

If the null hypothesis were proved to be false, a model would have to be designed, built and tested. This model would have to assist IT/IS learners in finding relevant information on the Internet, which could lead them to resources they would otherwise not have found. They should be able to do this in a way that is more efficient than unaided searching. According to Forcese and Richer, "A model is an imitation or an abstraction from reality that is intended to order and to simplify our view of that reality while still capturing its essential characteristics" (Forcese and Richer 1973:38). The proposed searching model would appear to be a search engine ("... imitation ... of reality"), which would hide the technicalities of search engines ("... simplify our view ..."), but still provide useful answers to users' queries ("... still capturing its essential characteristics"). The proposed model will fit this definition.

The problem of relevance judgement surfaces at this point. Mizzaro claimed that relevance judgements are affected by between 38 and 40 variables (Mizzaro 1998:303-304). Burgin noticed disagreement between authors and non-authors of queries about the relevance of answers to those queries (Burgin 1992:619). According to this author, there was a definitive relationship between subject knowledge and number of relevant documents as judged by an individual: the higher the level of subject knowledge, the fewer documents will be considered relevant. On the other hand, Courtois and Berry claimed that WWW search engines are easy to use since their results are based on relevancy searching (Courtois and Berry 1999:39). It is clear that confusion exists about how relevance of retrieved data is measured, and how reliable these measurements are.

Su used "... real questions and real end users' relevance judgements..." in a study to determine the adequacy of relevance for retrieval system evaluation. The author decided to follow this approach and allow the participants in this study to:

- define their own search topic within the broad limits of their study field,
- use these "...real questions..." to construct a query, and
- make their own "... real end users' relevance judgements..." on the answers they found (Su 1993:93).

Both the Cranfield tests and Lancaster's work on MEDLARS relied to an extent on a similar experimental approach as the one proposed for this study. During the Cranfield tests, for example, extensive searching was done on a defined set of documents, and decisions were taken on the relevancy of the results (Cleverdon, Mills and Keen 1966, Robertson 2001, Tonta 2001).

#### **4.1.4 POPULATION**

In line with the standard approach, the whole (population) from which the part (sample) had to be taken, was now defined (Forcese and Richer 1973:121).

The population was to be those who fall into the category of concern, viz. IT/IS learners. For the purposes of this study, a learner who is registered for at least one subject which contains a major IT/IS component would be part of the target population. Although this study was not limited to South African learners, budgetary constraints prescribed that they would form the bulk of the population. Universities and technikons have traditionally supplied the local market with IT/IS learners, via qualifications such as a B.Sc. (Computer Science), B.Comm., National Diploma in Information Technology and National Diploma in Financial Information Systems. As a result, local universities and technikons, plus similar international institutions would form the population for this study. Those institutions where a controller could be identified to assist would be included in the study. The controller would be necessary to arrange computer rooms, groups of learners, Internet access and other facilities ahead of time.

#### **4.1.5 SAMPLE**

According to Alreck and Settle, a correctly chosen sample would present an accurate extract of the population, without bias or error in the results (Alreck and Settle 1985:63). Kothari confirms that the choice of sample size is crucial. If too small, it may not achieve the objective of accurate measurement, and if too large, wastage of resources and huge costs may be incurred (Kothari 1991:212). Thus accuracy of the measurement must be balanced against cost and feasibility (Sapsford and Jupp 1996:28).

Furthermore, as a result of the complex racial composition of the inhabitants of the country, it was considered necessary to ensure that all easily definable race groups be represented in the sample (in alphabetical sequence: African, Asian, Coloured and White). These race categories are commonly used in South Africa as a means of determining to what extent the government policies of Equal Opportunity and Redress are being followed. This study does not focus on racial issues, and the inclusion of the race of participants was done merely to ensure that the results would be representative for the whole of South Africa. It was thus decided to draw as large a sample as possible from the population, to minimise sampling error.

Janes claimed that using a relatively small sample size can still provide useful answers, and provided a formula for the calculation of sample size, when the population size and other variables are known. The formula is:

$$n = 1 / [(1/N) + (e^2 / (Z^2 \times p \times (1-p)))] \text{ (Janes 2000:97,100).}$$

This formula to calculate sample size (n) was used as a control measure to determine the minimum sampling size (Janes 2000:100). It is based on population size (N), measure of error (e), measure of confidence (Z) and assumed proportion in population (p). A population size of 14 400 was obtained as follows: the estimated average number of IT/IS learners per institution was taken to be 400, which is a typical figure for the Cape Technikon. There are 21 universities and 15 technikons in South Africa, for a total of 36 institutions. Therefore:  $N = 400 \times 36 = 14\,400$ .

The values for e, Z and p were taken as recommended by Janes (e=0.04, Z=1.96, p=.05). His claim that changes in N have little effect on n was confirmed by the figures in the table below. This formula produced a value of n=576 (with N=14400) learners for this study.

N	n
5 000	536
14 400	576
15 000	577
25 000	586

**TABLE 4.1 – Interaction of Sample and Population size**

A further control measure was the e-mail based statements by 28 colleagues from higher education institutions from all over the world (mostly the USA, but also Argentina, Mexico, New-Zealand, Norway, Romania and Venezuela) stating the number of IT/IS learners in their institution. The average number of IT/IS learners from these messages was calculated to be 660.

A recalculation of N (population size) was done, using this new average of 660 instead of 400 as above, and using the same number of institutions of 36. This new calculation yielded an N of 23 760, which produced a new value for n of 585.

The author would attempt to involve at least 585 learners in the searching experiments.

#### **4.1.6 CONTROL**

To ensure reliable results, a measure of control had to be implemented. Both systematic and unsystematic factors were considered (Smit 1983:31). Systematic factors included the time of the proposed experiment, the measuring instrument, the instructions to the participants and the way the eventual solution to the problem would be implemented. Unsystematic factors included the research procedure and sampling.

#### **4.2 RESEARCH METHODS**

At this point the author had to decide how data about searching by IT/IS learners would be gathered, to be able to construct a model. The data would include: demographic features, choice of search engine, query specification, time spent searching and searching success or failure. The research method chosen should make this kind of data gathering possible, and should produce reliable, valid results. Above all, it should allow the author control over the situation in which searching takes place. This was especially true when the testing of the proposed model was to be done. Two experiments would have to be run to allow comparison of before and after results: the first one where the model was not applied, and the second one after having applied the model.

The author considered a number of research methods to be used for data gathering. A critique of these methods follows.

## **NON CONTACT METHODS**

### **4.2.1 POSTAL QUESTIONNAIRE**

Oppenheim defined a questionnaire as "... a set of questions, ... from more rigidly constructed scales or tests" (Oppenheim 1992:100), while Behr called it "... a document ... to be filled out by the respondent himself in his own time" (Behr 1988:155,156). Good estimated that this method has been used in more than 50% of the total research studies in education (Good 1963:271). It therefore appeared to be a possible starting point for this project. If this method had to be used, it would involve designing a questionnaire, which would be posted to participants. It would have to guide the participants to do an Internet searching experiment on their own.

Applebee and others used the mailed survey method in a study on Australian academics' use of the Internet, rather than an e-mailed survey. This was done not to discriminate against those academics who were not using the Internet. The response rate was 51.3% after two follow-ups (Applebee and others 2000:143). Majid and Abazova achieved a 63.3% response rate in a similar study on the academics of the International Islamic University of Malaysia (Majid and Abazova 1999:75).

However, a variety of problems were found with this method:

#### **4.2.1.1 Non-response**

Behr claimed that a response rate of less than 70% lacks validity (Behr 1988:162). A number of authors found the typical response figure to be substantially lower than this value. Gardner stated that a typical response rate is 15 – 50% (Gardner 1978:84). Weideman had a response rate of 49,0 % for a focused postal questionnaire (Weideman 1994:24) and Voorbij claimed 49.9% in a similarly focused postal questionnaire (Voorbij 1999:599). Marcella and Baxter had a 45.7% response rate for a UK-based mailed survey (Marcella and Baxter 1999:161).

Evans attempted to determine the identity and training of the authors of academic library home pages, and had a response rate of 54.9% to a mailed survey. An interesting finding of this study was that some library management personnel still considered the web as not being part of "...real' librarianship..." (Evans 1999:310,317).

One notable exception to these low figures is the results of a study done on health workers in Scotland. Farmer, Richardson and Lawton reported a high response rate of 91% for a questionnaire on improving access to information for nursing staff, while looking at the potential of the Internet (Farmer, Richardson and Lawton 1999:53).

#### 4.2.1.2 Questionable validity

Behr stated that the validity of postal questionnaire results was difficult to determine (Behr 1988:163). According to Gardner, items could be omitted, replies could be vague, and instructions could be confusing (Gardner 1978:84).

#### 4.2.1.3 Questionable reliability

Behr stated that the reliability of postal questionnaire results was low (Behr 1988:163), while Gardner referred to the "... incomplete sample resulting from questionnaires" (Gardner 1978:83).

#### 4.2.1.4 Respondent bias

Behr stated that respondents could misinterpret questions, resent interference by the questionnaire, or falsify answers (Behr 1988:163). These factors resulted in bias, compounding the problem of result reliability.

#### 4.2.1.5 Control

Alreck and Settle claimed that the environment of the respondent (time of day, location) when using this method may differ widely (Alreck and Settle 1985:44). Gardner referred to the fact that "There is no control over the situation..." and that "There is very little control over who answers the questions..." (Gardner 1978:85). This disadvantage of lack of control was considered to be the most serious of all problems discussed above.

It would be necessary to ensure that participants were from the identified target group, and would do the search experiment in a controlled, repeatable environment – this method appears not to guarantee that environment.

#### 4.2.1.6 Conclusion

According to the decision about sample size taken in 4.1.5 above, the non-response and lack of control problems disqualified postal questionnaires as a research method. Similarly, the potential validity problems indicated that this method could yield invalid results for the purpose of building a search model. As the proposed model cannot be based on unreliable data, this method was disqualified.

### 4.2.2 INTERNET QUESTIONNAIRE

An Internet-based questionnaire has the potential of delivering a high number of respondents, owing to the high number of Internet users – Strategis claimed that 100 million adults in the USA (more than half the population) use the Internet (Strategis 2000), while NUA figures estimated a world total of 407.1 million in November 2000 (NUA 2001).

However, it was scrapped since it would be impossible to establish the environment of participant (time spent on experiment, etc) or his association (i.e. is he/she an IT/IS learner?).

### 4.2.3 CASE STUDY

Both Behr and Smit claimed that a case study should be carried out on various attributes of one individual (Behr 1988:110, Smit 1983:47). Allan and Skinner described a variety of case studies, all concentrating on the behaviour of one small group of humans (Allan and Skinner 1991:191-200). Labovitz and Hagedorn claimed that the case study method is useful to describe a unit rather than for the testing of a hypothesis (Labovitz and Hagedorn 1981:48).

Clearly the size of the sample in the case study method would be inadequate to produce enough data to allow the author to design the proposed model.

## **CONTACT METHODS**

### **4.2.4 PERSONAL INTERVIEW**

Ample evidence exists that personal interviewing is an expensive research method (Alreck and Settle 1985:222, Dillman 1978:68, Gardner 1978:81, Hoinville and Jowell 1982:15, Oppenheim 1992:82).

Howcroft and Mitev decided to use a combination of the survey questionnaire and personal interview methods in an empirical study of Internet usage in UK medical practice management. They claimed that this combination delivered "...patterned diversity..." which cancels out some of the weaknesses of the individual methods. The response rate of the mailed questionnaire was, however, only 28% (Howcroft and Mitev 2000:172,173). Bruce used a structured interview method in a study on what people think when they search the Internet for information (Bruce 1999:189).

Had the personal interview method been chosen, each participant would have had to spend up to 180 supervised minutes to complete the first search experiment, receive exposure to the model, do the second experiment and record the results. This implied a large amount of time, and high travelling expenses for the author. Individual attention on a per person basis would be impossible if a large sample was desired. Since neither time nor financial support would be in abundant supply, this method was not considered further.

### **4.2.5 TELEPHONIC INTERVIEW**

This method suffers from similar drawbacks as personal interviews. Gardner mentioned the cost factor, time, lost production, interviewer training and the fact that this work is very tiring (Gardner 1978:81). Mohsin referred to the "no response" drawback, plus superficial information, respondent reluctance, lack of relationship between interviewer and interviewee and lack of checking (Mohsin 1984:171).

The lack of control is referred to by Alreck and Settle as "... the need for supervision", and other drawbacks mentioned include lack of eye contact, physical presence, trust and confidence (Alreck and Settle 1985:43).

Finally, the practical aspect of the proposed experiment implied that each respondent would have to be called twice: once to explain the details of the experiment, and again to obtain the results. This fact would double most elements of the cost and time factors. Furthermore, it would be impossible to expect the participant to understand and record the contents of one or two pages of information during a telephonic conversation. As a result this method was excluded as a possibility.

#### **4.2.6 LABORATORY EXPERIMENT**

Mohsin referred to the advantages of this method as being the high degree of control over variables (Mohsin 1984:145-147). However, the author considered the aim of large sample size, and realised that it would be impossible to convince large numbers of respondents from other institutions to travel to a fixed point for the experiment. It would also exclude learners from other parts of the country. Exorbitant costs would make this method unsuitable to implement.

#### **4.2.7 FIELD EXPERIMENT**

Petersen defined an experiment as "... a planned inquiry to discover new facts, or to confirm or deny the results of previous investigations" (Petersen 1985:5). Many of the results of previous work, as discussed in Chapter 2, needed to be confirmed in this study. Lovell and Lawson claimed that "A carefully controlled and rigorously conducted experiment is probably the most exact and refined method of research" (Lovell and Lawson 1970:82). Smit stated that field experiments provide the element of control expected from scientific research (Smit 1983:46). Mason and Bramble defined the experimental method as the "... systematic manipulation of experimental conditions in which extraneous influences are controlled ..." (Mason and Bramble 1989:54). Fidel and others claimed that the analysis of searcher behaviour in actual situations is a promising method for system design (Fidel and others 1999:36).

Huysamen confirmed the value of this kind of experiment done with groups, especially if it is done at their own "home location".

It was noted that some of its advantages include lower cost (since one person can administer it), and full control over the situation (since the administrator is present at every session) (Huysamen 1994:150,151).

Another author stressed the importance of the reliability of data collected, by stating that a valid measuring instrument used on different groups should produce the same observations (Mouton 1996:144). Leedy described the chosen method as follows: "Test the subject; do something to the subject; test the subject again to see if there is a difference" (Leedy 1985:210).

The author decided to use the field experimental method, based on the following points:

- Petersen, and Lovell and Watson's definition and description of the experimental method confirmed the methodology the author considered.
- Smit's mention of the element of control supported the scientific nature of the proposed fieldwork.
- Approval of the idea of testing searcher behaviour in a real-life situation by an eminent researcher (Fidel).
- Huysamen supported the concept of working with groups in a familiar environment.
- Finally, Leedy's description of the proposed methodology confirmed that the field experiment methodology was best suited to the circumstances.

A variation of the single-group experiment by Behr was to be used, where a single group at a time is subjected to different treatments and the resultant changes (if any) are measured (Behr 1988:116).

The elements of the proposed experiment would be:

Experimental group -

a group of IT/IS learners in a computer laboratory – the first measurement would be taken before the stimulus is applied.

Control group -

this would be the same group of learners as the experimental group, but the second measurement would be taken after the stimulus had been applied – where a learner was present on only one of the two occasions, his/her response would be classified as invalid.

Dependent variable -

the response of the experimental group in terms of their ability (or lack thereof) to find information on the Internet.

Independent variable -

the stimulus applied to the experimental group, being exposure to the use of the searching model.

Conditions -

these would include: searching time allowed, level and amount of instructions given, association of participants (i.e. are they IT/IS learners) and time of day (since it affects Internet traffic, and therefore response times) (Behr 1988:114, Mohsin 1984:130,131).

#### **4.3            PHASE 1 – DESIGN AND TEST INSTRUMENT: FORM 1**

The purpose of this first phase was to design and test a questionnaire. This questionnaire would be used in the field experiment to measure various factors about Internet searching done by learners. The questionnaire was called Form 1 (see Appendix A), and requested basic demographic details from the user. These would be used later to confirm that the participants in this study were not taken from a selected group only (i.e. only one gender, one race, a certain type of institution or a certain age group). The way in which resources were apportioned in the South African education system in the past, resulted in non-homogeneous groups of learners. Imbalances in resources between the various institutions could manifest in varying degrees of preparedness for independent study, hence the need to ensure wide enough coverage to include all combinations of genders, races and institutions.

The participants were then instructed to do three separate searches for any academic topic, and record the results. Before the actual searching and timing started, the concept and questionnaire were explained to the sample group.

However, no guidance was given on search engines, URLs, search methods, or any other relevant information on Internet searching.

A total time limit of 30 minutes for all three searches was set. Most of the institutions participating in the study offered 45 to 50 minute practical lectures. After subtracting time to log in, take roll call, explain what was to be done during the practical, etc, 30 minutes was considered to be a fair time left to do the actual work. Furthermore, personal experience suggested that a certain level of frustration set in if information eluded discovery on the Internet after approximately 30 minutes of searching.

Finally, due to the close proximity of most campus libraries, it was argued that a learner would consider it easier to find information in a library than to spend more than 30 minutes on a search.

This phase involved 105 learners taking part in five sessions in one South African educational institution.

SESSION NUMBER	CITY	INSTITUTION	NO. OF LEARNERS
1	Cape Town	Cape Technikon	29
2	Cape Town	Cape Technikon	23
3	Cape Town	Cape Technikon	27
4	Cape Town	Cape Technikon	13
5	Cape Town	Cape Technikon	13
<b>TOTAL</b>	<b>5</b>	<b>1</b>	<b>105</b>

**TABLE 4.2 – Phase 1 Session 1 - 5**

#### **4.4      PHASE 2 – INTERPRET RESULTS, MODIFY FORM 1 > FORM**

##### **2**

The information on the returned forms from Phase 1 was used to change the questionnaire to be a more accurate and useful measuring instrument. The following transpired from Phase 1:

##### **4.4.1      FOCUS**

The form lacked focus, since it did not specify exactly what the experiment was about (verbal feedback).

##### **4.4.2      FEEL**

The formatting of the document did not include sufficient differentiation in the use of font size, line thickness and other features, which made the document difficult to read and pinpoint important parts (verbal feedback).

##### **4.4.3      THREE SEARCHES**

The requirement that each participant does three separate searches and record all results in 30 minutes appeared to be too stringent. Fifty out of the 105 participants noted down a variety of comments, all pointing to this fact. Some comments were noted at the bottom of Form 1:

Participant number 15:	"Search took very long"
Participant number 38:	"Speed and huge data load"
Participant number 60:	"Ran out of time"
Participant number 94:	"Takes too long on each subject"
Participant number 105:	"Didn't finish".

##### **4.4.4      FEEDBACK**

There was no space for the participant to supply written comments on the experiment (verbal feedback).

##### **4.4.5      FALSE YES**

It was noted that some participants gave a "Yes" answer to indicate that they did find the information, while the information on their form indicated it to be an invalid answer.

## Some examples:

- Participant number 6: The participant did not specify what he/she was searching for.
- Participant number 15: The participant's specification was far too wide to be useful ("Information Technology"), and it would be almost impossible not to find some information on it.
- Participant number 72: The participant answered "Yes" without completing any other detail on the form.
- Participant number 90: The participant answered "Yes" to all three searches, claimed to have spent 53 minutes searching (all experiments were timed by the author personally to exactly 30 minutes), and claimed to have visited 18 sites covering three diverse topics (in only 30 minutes).
- Participant number 95: The participant did not use a search engine, visited "all" sites found, and spent only eight minutes on the search.

This phenomenon was attributed to the fact that achieving success in Internet searching in a group situation could enhance the standing of the participant in the group.

#### 4.4.6 SUCCESS RATE

The results from this first group indicate that 64.9% of the participants did not find the information they were looking for, and only 35.1% did (see Appendix J, J.1.8 – the number of "Unknown" answers were ignored). At this stage it appeared that  $H_0$  was false and that  $H_1$  would have to be proved to be true.

Based on these results, the following changes were made to Form 1:

**4.4.7 HEADING**

A heading was added to state clearly what the experiment was about.

**4.4.8 READABILITY**

The outlines of boxes were made thicker to make them more visible, and different font sizes were used for clarity.

**4.4.9 THREE SEARCHES**

The requirement that each participant does three separate searches was omitted – only one search had to be done.

**4.4.10 FEEDBACK**

An extra box was added where participants could add their comments.

**4.4.11 IDENTIFICATION**

The "Initials" and "Surname" boxes were removed, thereby removing some of the motivation for giving false "Yes" answers.

The resultant Form 2 is attached as Appendix B.

**4.5 PHASE 3 – TEST FORM 2 AS INSTRUMENT**

Form 2 was now used to repeat the test in one session at a South African university, involving 19 learners.

SESSION NUMBER	CITY	INSTITUTION	NO. OF LEARNERS
6	Stellenbosch	University of Stellenbosch	19
<b>TOTAL</b>	<b>1</b>	<b>1</b>	<b>19</b>

**TABLE 4.3 – Phase 3 Session 6**

#### **4.6      PHASE 4 – INTERPRET RESULTS, MODIFY FORM 2 > FORM 3**

The feedback from Phase 3 was used to change the questionnaire again. The following points became evident after Phase 3:

##### **4.6.1      INSTRUCTIONS**

Simpler and fewer instructions were required (verbal feedback).

##### **4.6.2      FLOW**

A section was needed which would lead the learner through a logical process from defining a problem experienced with his/her current study materials, to finding the solution on the Internet.

##### **4.6.3      METHODOLOGY**

Participants also requested that more space be made available to record details about the method they followed and the results.

##### **4.6.4      RESULTS**

The results from this second group indicated a 50/50 split between the "Yes" and "No" answers (see Appendix J, Section J.2.10 – the number of "Unknown" answers were ignored). A decision about accepting the null or alternative hypothesis was thus delayed until more data became available.

Based on these results, the following changes were made to Form 2:

##### **4.6.5      INSTRUCTIONS**

The instructions throughout the form were simplified and shortened.

##### **4.6.6      PROBLEM IDENTIFICATION**

The question about prior exposure to the Internet and searching was broadened to elicit more detail from the participant. The learner was also "forced" to specify a search which would (the author hoped) solve the problem just defined.

##### **4.6.7      METHODOLOGY**

A separate page was added with more space to record detail about the method (or lack thereof) followed when searching.

The resultant Form 3 is attached as Appendix C. Most of the subsequent experiments by learners were done using Form 3 (and later Form 4) for the unassisted section of the searching experiment. A small block for identification purposes was added to Form 3, changing it into Form 4 – otherwise the two are identical.

#### **4.7**            **SUMMARY**

The stage had now been set for the major section of the field experiments. The measuring instrument had been designed, tested, improved, re-tested and was ready for implementation.

# **CHAPTER 5**

## **CONDUCT OF THE STUDY**

### **5.1 INTRODUCTION**

A brief summary of the experimental work done up to this point follows.

#### **5.1.1 HYPOTHESIS**

A hypothesis was postulated, stating that learners either could or could not find academic information on the Internet without assistance.

#### **5.1.2 POPULATION / METHODOLOGY**

A population size was estimated and a sample size calculated, and the field experiment research method was chosen to gather data.

#### **5.1.3 INSTRUMENT**

A measuring instrument was designed and tested on 105 learners.

#### **5.1.4 REFINEMENT AND TEST**

The results from these learners were used to improve the instrument, and it was then tested again on 19 other learners.

#### **5.1.5 REFINEMENT**

These results were then again used to refine the test instrument.

#### **5.1.6 IMPLEMENTATION**

The next step was to use this instrument on a large scale to determine how learners search for information, what their success rate is and what the reasons are for their success and failure.

## 5.2 PHASE 5 – USE FORM 3

A country-wide tour was undertaken to use Form 3 as a measurement of learner's searching methods, and as an instrument to attempt to determine the reasons why some are successful in searching and others not.

The same searching experiment at each institution was done as in Phase 3 (see Section 4.5). A group of learners would use Form 3 to do a search for one piece of academic information in 30 minutes.

To ensure that the results of this study were consistent for IT/IS learners from all races/genders/higher education institution types (see Section 4.1.4), an attempt was made to include as many of these learners as possible in the sample. Colleagues were identified at universities and technikons covering the whole of South Africa – from Cape Town in the far south, to Bloemfontein, Welkom and Port Elizabeth in the centre to Johannesburg and Pretoria in the north, and Durban and Pietermaritzburg in the east. These colleagues, or controllers, played an important part in the preparation of the actual experiments in terms of computer laboratories, learner availability, hardware and software requirements, etc.

Phase 5 involved 293 learners divided into 15 sessions at 12 South African institutions, spread over seven cities.

SESSION NO.	CITY	INSTITUTION	NO. OF LEARNERS
7	Durban	University of Durban – Westville	19
8	Durban	University of Durban – Westville	16
9	Durban	University of Durban – Westville	5
10	Pietermaritzburg	University of Natal (PMB)	20
11	Durban	Technikon Natal	7
12	Pretoria	Vista University – Mamelodi	16
13	Pretoria	Vista University – Mamelodi	8
14	Johannesburg	Rand Afrikaans University	49
15	Johannesburg	Technikon Witwatersrand	15
16	Port Elizabeth	Vista University – PE	28
17	Port Elizabeth	PE Technikon	19
18	Bloemfontein	Technikon OFS	24
19	Bloemfontein	Vista University – BFN	38
20	Bloemfontein	University of the Free State	5
21	Cape Town	Cape Technikon	24
<b>TOTAL</b>	<b>15</b>	<b>7</b>	<b>12</b>
			<b>293</b>

<b>TABLE 5.1 – Phase 5 Session 7 - 21</b>
---

### 5.3 PHASE 6 – INTERPRET RESULTS, DECISION ON MODEL

The results available at this point were considered substantial enough to take a decision on the hypotheses postulated earlier (see Section 4.1.2). Enough detail was also available from summarising results to allow the author to identify some of the reasons why some learners were successful and others not.

The Yes/No results from the three groups involved in the experiments so far are summarised below.

PHASE	YES	NO	% NO
1	34	63	64.9
3	8	8	50
5	90	135	60
<b>TOTAL</b>	<b>132</b>	<b>206</b>	<b>60.9</b>

**TABLE 5.2 – Yes/No answers Phase 1, 3, 5**

With an overall average of 60.9% of learners (206 out of 338) claiming that they could not find one piece of academic information in 30 minutes unaided, it was clear that the creation of a model to assist learners was justified. Hypothesis  $H_0$  was thus proved false, and hypothesis  $H_1$  would have to be proved to be true.

The author now had to decide on the format of the model(s). The options were to use paper forms or a web-based model (appearing to be a screen form to be filled in). It was decided that initial model(s) should be paper-based forms. If the model was available on the web at this stage, some learners could access it before the first experiment, thereby rendering their results useless without the author being aware of it. The paper forms filled in by the learners could simply be used as is to provide the required feedback. However, the final version of the model would have to be web-based, since that is the environment in which it should be used on a daily basis.

Furthermore, users should be able to access it without the author being present, and it would also allow changes to be made easily if it were published on the web.

The following problems were identified after reading the forms of the 293 participants of Phase 5:

### **5.3.1 SEARCH ENGINES**

Learners did not know which search engine(s) to use. Since hundreds of search engines are available to the user, and since a number of these have been well known over the past years, an even spread across these programs was expected. Instead, a disproportionately high percentage (50.9%) of learners chose Yahoo!, with the second and third choices totalling only 9.8% and 9.3% respectively (see Appendix J, Section J.3.8). Other researchers do not agree that Yahoo! is always the best choice to make: Hock investigated 13 search engines “.. which the serious searcher is likely to use at least occasionally” (Hock 1999:25). Yahoo! was not listed. In an article which keeps the reader up to date on search engine developments, Notess discusses seven commonly used search engines and new developments on their sites – again Yahoo! was omitted (Notess 1999a:17).

Possible reasons for the choice of the Yahoo! search engine could be its youthful image, or by participants copying the choice of their neighbours. Lack of knowledge about the choice to make was thus noted, and Model 1 would have to address it.

### **5.3.2 STRATEGY**

Once a search engine was chosen, a lack of insight into searching strategy was evident. Often single words or two-word phrases were submitted without any operators. Not even the most basic elements of correct search engine coding were applied, including correct spelling, use of capital letters and correct use of phrases. It appeared as if many of the participants did not know exactly what they were looking for.

In an attempt to classify these errors into categories, the forms of "invalid" participants (invalid for the purpose of defining errors) were first filtered out: those whose forms had any one or more of the following features:

- answered Yes to the "Did you find it" question,
- used searching operators correctly,
- used an invalid search engine (including simply stating the URL of the site containing the answer),
- no query was shown, and
- used directories instead of keyword searching.

It was found to be impossible to classify every remaining error into detailed classes, since some of the errors would depend on the syntax of the search engine chosen. It was also found that more than one mistake was made in one query, making it difficult to classify. However, it was possible to generalise and identify three broad types of errors:

- Type 1: incorrect use of phrases, including phrase too long, incorrect use of quotes (omitting them when the search engine requires it), and the use of terms which are too general in the phrase (for example: programming tutorial, system analysis, data communications).
- Type 2: the use of only a single, general word as query (for example: computer, education, programming).
- Type 3: incorrect use of operators (inserting a space between the operator and the keyword, for example: "+ Colnago + frame" instead of "+Colnago +frame").

The frequency at which these errors appeared is listed below.

TYPE	FREQUENCY
1	47
2	32
3	9

**TABLE 5.3 – Phase 6 Error Frequency**

In the final version of the model, these errors would be considered in the design.

### **5.3.3 SPEED**

Complaints about waiting time, slow downloading and time-out errors were still noted. Some participants actually abandoned their search well before the allotted 30 minutes had expired as a result of “speed” problems. Although no control over outside factors (bandwidth, server specifications, cache presence) was possible, it was considered necessary to address this problem in Model 1.

### **5.3.4 CHOICE OF SEARCH ENGINE**

The first problem (search engine choice – see Section 5.3.1) would be solved in Model 1 by giving the user a choice of search engines with their URLs. A decision thus had to be taken on which search engine programs to include in Model 1. A combination of the following three sources were used to aid in making this decision: search engines used by learners during prior experiments, those mentioned in research literature and those in use by the computing community. The resultant list was rounded off to 20 search engines, and is summarised in the table below. Learner choices (see Appendix J, Table J.1.5, J.2.7, J.3.8) are indicated in the USER column, and the other two combined are listed in the LITERATURE column (Habib & Balliot 1999, Sullivan 1999b).

USER	LITERATURE
	All-in-One
AltaVista	
	Ask Jeeves
	Dogpile
	Excite
Goto	
Hotbot	
Infoseek	
	Looksmart
Lycos	
	Magellan
	Mamma
Metacrawler	
	Metasearch
Netscape Search	
	Northern Light
	Profusion
	Savvysearch
	Snap
Yahoo!	

**TABLE 5.4 – Chosen Search Engines**

### 5.3.5 OPERATORS

The second problem (lack of strategy – see Section 5.3.2) would be addressed in two ways. First, the user would be requested to write out his/her search specification in one full English sentence. It was assumed that this procedure would force the user to think about the concept he/she was searching for.

Secondly, he/she would work through a series of guidelines as to how he/she could build up a query, based on this specification, which would be more effective than just one or more words. An inspection was made of the help menus of the 20 search engines listed in Table 5.4 above.

The most common search operators were identified, in terms of the number of search engines supporting their use: inclusion (+), exclusion (-), phrases (""), range limiting by domain (DOMAIN:...), range limiting by URL (URL:...), range limiting by webpage title (TITLE:...), and the wildcard (\*). Each operator was introduced in a single line statement, and supported by a series of worked out examples to illustrate its use (See Appendix F, Model 1, box 10).

### **5.3.6 SPEED**

The third problem (speed - see Section 5.3.3) was addressed in two ways. First the user was lead via two questions (See Appendix F, Model 1 Box 4 and Box 5) to sleeping servers: mirrored sites of established search engines in other countries. Depending on the user's time of day and location with reference to the GMT line, searching time could be decreased by executing a search on a server in a country where the local time is between 23h00 and 07h00. Secondly, guidance was offered on using the multiprogramming feature of Windows to run more than one search engine concurrently. This was done as a result of the suggestions made in research literature (see Section 2.5.7) that more than one database should be consulted when searching on the Internet.

If the user could load, say, four different search engines and start them searching one after the other, it would create slightly more network traffic with a resultant slowdown of response. However, at the same time it would raise the chances of obtaining an answer faster owing to some programs responding faster than others. At the same time it would "keep the user busy", as he/she would continuously switch between the windows, waiting for the first program to produce answers. This "being busy" would alleviate the negative perception of "slow computers" and hopefully urge more users to keep on trying until the time ran out.

More importantly, it would increase the chances of receiving relevant answers, since multiple databases are being queried (Chun 1999:137, Garman 1999:75, Jacso 1999a:227, Jacso 1999b:99, Mickey 1999:79, Williams 1996:246).

The resultant Model 1 was designed on paper – see Appendix F.

#### **5.4 PHASE 7 – TEST FORM 3 WITH MODEL 1**

The first full experiment with three components was then launched: a repeat of the standard first experiment using Form 3 (as in Phase 5), followed by a training session of 90 minutes, followed by a second experiment using Model 1 on the same group of learners.

Phase 7 was conducted in three sessions at two different institutions, in two cities and two countries (Slovenia and South Africa). A total of 68 learners were involved.

SESSION NUMBER	CITY	INSTITUTION	NO. OF LEARNERS
22	Kranj (Slovenia)	University of Maribor	18
23	Kranj (Slovenia)	University of Maribor	26
24	Cape Town	Cape Technikon	24
<b>TOTAL</b>	<b>3</b>	<b>2</b>	<b>68</b>

**TABLE 5.5 – Phase 7 Session 22 - 24**

#### **5.5 PHASE 8 – INTERPRET RESULTS, MODIFY MODEL 1 > MODEL 2**

The following was noted after reading the forms of the 68 participants of Phase 7, giving rise to the decisions mentioned:

##### **5.5.1 COUNTRY DEPENDENCE/SLEEPING SERVERS**

Only four participants chose the country dependence option.

The reason for the lack of interest in this option was considered to be the fact that academic topics in the IT field are seldom country related, as opposed to fields of study like History, Economics, Religion, etc. Only seven participants chose the sleeping servers option – possible reasons could be unreliability of smaller servers in less well-known countries, and more restricted coverage offered by smaller indices compared to the index of the main server.

It was decided to omit both these options, which would have the added advantage of simplifying the layout of Model 2.

### 5.5.2. QUERY

Few participants managed to build up a query with complexity beyond the very basic few words or maybe a phrase and exclusion operator, even though the author offered many examples using a variety of operators during the training session (see Appendix K – Examples). In Model 2 even more examples would be given to assist the user in constructing a reasonable query. In an attempt to make the form even simpler in appearance, the instructions were further simplified, and a more “user-friendly” font was used.

Model 1 was redesigned and called Model 2 – see Appendix G.

### 5.6 PHASE 9 – TEST FORM 3 WITH MODEL 2

Phase 9 involved 22 learners from two universities in the USA, spread over three sessions and two cities.

SESSION NUMBER	CITY	INSTITUTION	NO. OF LEARNERS
25	Dallas (USA)	University of Dallas	6
26	Dallas (USA)	University of Dallas	13
27	New York (USA)	Columbia University	3
<b>TOTAL</b>	<b>3</b>	<b>2</b>	<b>22</b>

**TABLE 5.6 – Phase 9 Session 25 – 27**

## **5.7            PHASE 10 – INTERPRET RESULTS, MODIFY FORM 3 > FORM 4, MODEL 2 > MODEL 3**

Although the results of testing with Model 2 were satisfactory, two major problems were noted.

### **5.7.1            QUERY**

Most learners were still unable to compile their own search queries, even after having been shown a number of complete examples (see Appendix K: Examples). It was decided that the last version of the model would have to extract enough information from the user, and it would have to construct the query based on that automatically. In this way the user would not be required to learn or apply the syntax of any search engine.

This decision was in line with the approach followed by other researchers regarding the automatic generation of search queries (Chen and others 1998:582,600,601, Frants and Shapiro 1991:16, Gauch and Smith 1993:124).

### **5.7.2            SEARCH ENGINE CHOICE**

The large choice of search engines also appeared to be confusing. Many users insisted on using only one of them, which implied that the parallelism of Windows was also not utilised to decrease waiting time. At the same time, the positive effect of querying multiple databases was lost. Model 3 would have to give the user fewer options of search engines, and make it even more enticing to use Windows in such a way that waiting time was minimized and more relevant answers could be obtained

Up to this point, the various sessions involved small enough numbers of learners to enable the author to use handwriting matching in marrying Form 3 of a learner before the training session, to his/her Model 2 form after the training session. However, with larger groups expected in the near future, a solution had to be found for the problem described below.

During the first session, the learners would complete and hand in Form 3, without any information which could identify them personally (surname, student number, etc). After the training session, they would then use the Model 2 form, and submit that, also without any personal information noted down. Depending on the institution, these sessions were sometimes done on different days. As a result of the potential problem of mismatching of the two forms, some device was necessary to enable proper matching. A small block was added in the top right hand corner of Form 3 (changing it to Form 4) and Model 2, where the learner could enter a random personal code, as long as the same code was used on both occasions.

Model 3 would be web-based, which has the following advantages:

- It would be available to any user anywhere in the world without the author's intervention.
- It would result in a natural flow, from typing in searching detail, to the calling of search engines, to the interpretation of answers on the screen.
- It would be easy to change later, while still being accessible to past participants as well as future learners.

The author named Model 3 FOIOTI, for Finder Of Information On The Internet. FOIOTI is available at <http://www.mwe.co.za>.

The purpose of FOIOTI is not to retrieve the information from the web – most search engines do that properly if given a complete specification. FOIOTI's purpose is to:

- extract enough information from the user to be able to construct a complete specification,
- construct the specification, and
- submit it in the correct format to the selected search engine(s).

This concept is in line with early studies where an algorithm was developed to construct a Boolean query (Frants and Shapiro 1991), and an expert system was developed to reformulate queries (Gauch and Smith 1993).

The next step was to determine which search engines should be included, and what methodology should be used in this model.

#### 5.7.2.1 SEARCH ENGINE CHOICE

A weighting system would be used to combine the 20 search engines identified earlier (see Section 5.3.4) with those chosen by the learners of Phase 5, 7 and 9 (session 7 up to 27). The popular search engines amongst the learners were considered important to include in the line-up for choosing those to be included in Model 3. Since most of the well-known, reliable search engines (AltaVista, Infoseek, Hotbot, etc) were represented in the group, it was considered a reliable sample from the population of all search engines. Furthermore, if they were popular among the learners, they would be popular as an option to use in the model.

Only one half of the search engines identified in Section 5.3.4 were selected on a scale of preference by the learners – the rest were chosen from research literature. As a result, a score of 10.5 points was allocated to all of them (10.5 is the average of  $20 + 19 + 18 + \dots + 2 + 1$ ). These scores are based on the assumption that scores between 1 and 20 would have been allocated to them if they were selected on preference, and is called Score 1 (see Table 5.7).

Secondly, all the search engines chosen by the learners of session 7 to 27 received a score, ranked according to the number of learners choosing them (column marked # USERS 2 in Table 5.7). Since there are 22 search engines in this group, the search engine being chosen by the most learners received a score of 22, then 21, etc. Those with an identical number of votes received identical scores. This figure was called Score 2 in Table 5.7. The last column indicates the sum of Score 1 and Score 2, where a higher figure indicates a better choice of search engine.

The results of this scoring system are given in Table 5.7.

SEARCH ENGINE	SCORE 1	# USERS 2	SCORE 2	SC.1 + SC.2
Aardvark	0	1	1	1
All-in-One	10.5	0	0	10.5
AltaVista	10.5	45	21	31.5
Ananzi	0	10	18	18
Ask Jeeves	10.5	1	1	11.5
Dogpile	10.5	0	0	10.5
Excite	10.5	8	16	26.5
Google	0	2	9	9
Goto	10.5	1	1	11.5
Hotbot	10.5	8	16	26.5
IBM	0	1	1	1
Infoseek	10.5	20	19	29.5
Looksmart	10.5	0	0	10.5
Lycos	10.5	6	15	25.5
Magellan	10.5	0	0	10.5
Mamma	10.5	1	1	11.5
Metacrawler	10.5	2	9	19.5
Metasearch	10.5	0	0	10.5
Microsoft	0	3	12	12
Netscape Search	10.5	2	9	19.5
Northern Light	10.5	0	0	10.5
Profusion	10.5	0	0	10.5
Savvysearch	10.5	0	0	10.5
Search	0	5	14	14
Snap	10.5	1	1	11.5
Webcrawler	0	26	20	20
Webferret	0	1	1	1
Webopedia	0	1	1	1
Yahoo!	10.5	153	22	32.5
Zazoo	0	3	12	12

**TABLE 5.7 – Search Engine Choice**

From Table 5.7 above, all search engines with a total (Score 1 + Score 2) of 20 and higher were chosen to be subjected to further tests. The search engines identified this way are listed in Table 5.8.

SEARCH ENGINE	TOTAL
Yahoo!	32.5
AltaVista	31.5
Infoseek	29.5
Excite	26.5
Hotbot	26.5
Lycos	25.5
Webcrawler	20

**TABLE 5.8 – Highest Search Engine Scores**

#### 5.7.2.2

#### SEARCHING METHODOLOGY

A summary was made of the methodology used by all those learners who did not find any information without a model, but did find it by using the model of Phase 7 and 9 (where the model was used). Since these learners used the operators suggested in the model to achieve success, it was assumed that these operators have a higher probability of generating successful results for other searchers as well. Their queries were inspected, and the operators used by these successful searchers are summarized below.

OPERATOR	# OF SEARCHERS
+ (Inclusion)	28
" " (Phrase)	14
* (Stemming)	4
- (Exclusion)	3
AND (Logical)	1
DOMAIN: (Field)	1
URL: (Field)	1

**TABLE 5.9 – Operators used by Searchers**

These findings confirmed those of Spink, Bateman and Jansen as well as Voorbij, who stated that searchers seem to favour simple operators and avoid more complex ones (Spink, Bateman and Jansen 1999:123,125, Voorbij 1999:603). It was decided to do further tests on these operators, to determine which one(s) produced the most relevant results when used in combination with the search engines listed in Table 5.8.

### 5.7.2.3 SEARCH ENGINE TESTS

The next step was to combine the use of the search engines identified in Table 5.8 above, with the operators identified in Table 5.9 above. Since this research project is aimed at an audience of real-world users, a real-world testing scenario was preferred above a theoretical one. Over 300 detailed Internet searches would be done by the author, the results recorded, and a final decision about the operation of Model 3 would be taken. The searches below start with a REQUEST each, which is based on a perceived information need (PIN) of a typical IT/IS learner (Mizzaro 1998:306). A variety of QUERIES were then to be constructed, based on the operators identified in Table 5.9 above. Half of the queries under REQUEST 1 to REQUEST 20 below were constructed from the popular operators (the first two rows of Table 5.9: Inclusion and Phrases), and the other half from the least popular operators (the last three rows of Table 5.9: Logical and Field).

Rows number three and four (Stemming and Exclusion) were considered useful mostly on repeated successive searches on the same topic. For example, a user does a search on Intel processors, using the query: "Intel" +processors. Only after receiving some incomplete answers, he/she realizes that the last "s" in the query could filter out wanted answers, then changes the query to "Intel" +processor\*. Similarly, a search for password rules in Windows NT using the query: Windows AND NT AND password could produce superfluous answers including all other Windows versions. Afterwards the user might change the query to: Windows AND NT AND password -98 -95 -3.\* to exclude unwanted version numbers. For this reason these two operators were excluded from the experiments.

Although the first ten queries were generated by the author, each one's topic was based on query topics of some of the participants up to this point. The author constructed each query to be generic in nature, and did not consider the exact syntax of any specific search engine. This was done to mimic the final version of the model, which would treat the components of the query as submitted by the user the same. Each query would be submitted to each one of the search engines identified in Table 5.8 above.

The first 10 results of each of the searches would be inspected in each case, since various authors noted that most users view only the first few results (Courtois and Berry 1999:43, Notess 1999c:84). A decision would then be taken about whether or not a relevant answer to the query was found amongst these ten websites. The author's own relevance judgement was considered reliable enough to make this judgement, since all the REQUESTS are in his field of knowledge.

Details about the interpretation of the next 20 tables and QUERIES follow.

- The first row in each table contains an abbreviation for each search engine listed in Table 5.8 above.
- The first column in each table lists the queries described directly above that table.

- The numbers in the other columns indicate the number of answers received. A lowercase t indicates a time-out error given by the browser, with no figure as a result. A lowercase x indicates an error message from the search engine concerning the query syntax, with no figure as a result. The figures in bold include at least one relevant site under the first 10 (i.e. success), as judged by the author.
- Queries based on the Inclusion and Phrase operators are:  
Q1.3, 1.5, 2.3, 2.5, 3.3, 3.5, 4.3, 4.5, 5.3, 5.5, 6.1, 6.3, 7.1, 7.3, 8.1, 8.3, 9.1, 9.3, 10.1, 10.3 plus all queries from 11 to 20 both included.
- Queries based on Logical and Field operators are all those not listed above.

### **REQUEST 1**

**I need to find out how to set up a query based on a table using Microsoft Access.**

#### QUERY 1.1

TITLE:Microsoft AND table +query AND "Access"

#### QUERY 1.2

TITLE:Microsoft OR table +query OR "Access"

#### QUERY 1.3

query +table +"Microsoft" +"Access" +tutorial

#### QUERY 1.4

(TITLE:Access OR URL:Access) AND (+query +table +tutorial)

#### QUERY 1.5

+"Access" +tutorial +table +query

	AV	IS	YA	LY	HB	WC	EX
Q1.1	2 798 960	92	0	0	0	t	0
Q1.2	1 136 672	822 140	59	13	82	0	0
Q1.3	0	272	8	0	4 550	149	4 538
Q1.4	20	0	143	0	0	t	t
Q1.5	0	0	8	0	6 600	79	3 098

**TABLE 5.10 – Results from Request 1**

## REQUEST 2

I need a reference on the assembler instruction set for Intel microprocessors.

### QUERY 2.1

URL:Assembler AND instruction +reference AND "Intel"

### QUERY 2.2

URL:Assembler OR instruction +reference OR "Intel"

### QUERY 2.3

"Intel" +assembler +instruction +reference +"808\*\*"

### QUERY 2.4

(TITLE:Assembler OR URL:Assembler) AND (+reference +instruction +set +Intel)

### QUERY 2.5

+"Intel" + assembler +instruction +reference

	AV	IS	YA	LY	HB	WC	EX
Q2.1	3	0	0	0	0	0	0
Q2.2	1	250 934	0	9	0	0	0
Q2.3	0	0	71	0	1 100	12	1 125
Q2.4	12	0	118	0	0	t	t
Q2.5	1	0	3	0	1 670	15	872

**TABLE 5.11 – Results from Request 2**

### REQUEST 3

I need to find out what the V.90 modem specification means.

#### QUERY 3.1

TITLE: Modem AND modem +specification AND "V.90"

#### QUERY 3.2

TITLE: Modem OR modem +specification OR "V.90"

#### QUERY 3.3

modem +"V.90" +specification

#### QUERY 3.4

(TITLE:V.90 OR URL:V.90) AND (+modem +specification)

#### QUERY 3.5

+"V.90" +modem +specification

	AV	IS	YA	LY	HB	WC	EX
Q3.1	4 810 669	0	0	0	10	0	0
Q3.2	1 425 335	22 808	0	10	82	t	t
Q3.3	736 375	98 338	2	10	82	x	x
Q3.4	0	0	23	0	0	x	x
Q3.5	56	28	2	0	1 360	48	2 191

**TABLE 5.12 – Results from Request 3**

**REQUEST 4**

I need a description of the operation of the Melissa computer virus.

**QUERY 4.1**

TITLE:Virus AND Melissa +description AND "Melissa"

**QUERY 4.2**

TITLE:Virus OR Melissa +description OR "Melissa"

**QUERY 4.3**

"Melissa" +computer +virus +description

**QUERY 4.4**

(TITLE:Virus OR URL:Virus) AND (+Melissa +computer +virus +description)

**QUERY 4.5**

+"Melissa" +computer +virus +description

	AV	IS	YA	LY	HB	WC	EX
Q4.1	720 909	0	1	0	290	x	x
Q4.2	254 893	95 969	1	20	82	x	x
Q4.3	26	1 539	418	0	1 040	308	9 438
Q4.4	374	0	9	0	0	x	x
Q4.5	0	18	11	0	1 040	0	331

**TABLE 5.13 – Results from Request 4**

**REQUEST 5**

I need to find out what the difference is between Active pages and CGI scripts in webpage design.

**QUERY 5.1**

TITLE:Web OR Active +pages +CGI OR "Webpage design"

**QUERY 5.2**

URL:Web OR Active +pages +CGI OR "Webpage design"

QUERY 5.3

active +pages +CGI +scripts

QUERY 5.4

(TITLE:Web OR URL:Web) AND (+active +pages +CGI +scripts)

QUERY 5.5

+active +pages +CGI +scripts +difference

	AV	IS	YA	LY	HB	WC	EX
<b>Q5.1</b>	<b>3 214 989</b>	<b>1 153 236</b>	<b>8</b>	<b>5 000</b>	<b>101</b>	<b>x</b>	<b>x</b>
<b>Q5.2</b>	<b>2 392</b>	<b>243 017</b>	<b>8</b>	<b>20</b>	<b>0</b>	<b>x</b>	<b>x</b>
<b>Q5.3</b>	<b>0</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>35 050</b>	<b>1 345</b>	<b>2 530</b>
<b>Q5.4</b>	<b>846</b>	<b>0</b>	<b>9</b>	<b>0</b>	<b>0</b>	<b>x</b>	<b>x</b>
<b>Q5.5</b>	<b>0</b>	<b>0</b>	<b>9</b>	<b>0</b>	<b>6 730</b>	<b>23</b>	<b>828</b>

<b>TABLE 5.14 – Results from Request 5</b>
--

REQUEST 6

I want to find out how to embed a Java script in an HTML document.

QUERY 6.1

"Java" +script +embed +HTML

QUERY 6.2

(TITLE:Java OR URL:Java) AND (+embed +script +HTML)

QUERY 6.3

+"Java" +script +HTML +embed

	AV	IS	YA	LY	HB	WC	EX
Q6.1	0	3	25	0	9 150	73	2 780
Q6.2	185	0	12	0	0	x	x
Q6.3	0	0	112	0	10 480	38	1 662

**TABLE 5.15 – Results from Request 6**

### REQUEST 7

I need to find out if Windows NT Server V4.0 makes use of semaphores to achieve mutual exclusion.

#### QUERY 7.1

"Windows NT" +"Server" +"V4.0" +semaphores

#### QUERY 7.2

(TITLE:Windows OR URL:Windows) AND (+Server +4.0 +semaphores)

#### QUERY 7.3

+"Windows NT Server" +4.0 +semaphores

	AV	IS	YA	LY	HB	WC	EX
Q7.1	21	1	0	1	32	0	100
Q7.2	107	0	75	0	0	x	x
Q7.3	0	0	2	0	0	0	25

**TABLE 5.16 – Results from Request 7**

### REQUEST 8

I need to understand the basic concepts of EDI.

#### QUERY 8.1

"EDI" +basic +concepts

QUERY 8.2

(TITLE:EDI OR URL:EDI) AND (+basic +concepts)

QUERY 8.3

+"EDI" +basic +concepts

	AV	IS	YA	LY	HB	WC	EX
Q8.1	447 155	32 080	2 261	0	4 580	2 592	29 710
Q8.2	36	3	15	0	0	x	x
Q8.3	2 295	179	2	0	4 580	54	1 136

TABLE 5.17 – Results from Request 8

REQUEST 9

I need a complete definition of Object-oriented Analysis and Design.

QUERY 9.1

definition +object +oriented +analysis +design

QUERY 9.2

(TITLE:Object OR URL:Object) AND (+definition +analysis +design)

QUERY 9.3

+"Object" +oriented +analysis +design

	AV	IS	YA	LY	HB	WC	EX
Q9.1	0	2	1	0	26 060	591	2 300
Q9.2	0	0	2	0	0	x	x
Q9.3	13 830	613	460	0	30 910	0	2 300

TABLE 5.18 – Results from Request 9

**REQUEST 10**

I need to find some free downloadable tutorials on the general use of SQL.

**QUERY 10.1**

"SQL" +tutorial +download

**QUERY 10.2**

(TITLE:SQL OR URL:SQL) AND (+query +use)

**QUERY 10.3**

+"SQL" +tutorial +download

	AV	IS	YA	LY	HB	WC	EX
<b>Q10.1</b>	6 285	139	26	0	<b>6 460</b>	1 920	4 520
<b>Q10.2</b>	0	<b>13</b>	143	0	0	x	x
<b>Q10.3</b>	<b>24</b>	0	4	0	<b>6 460</b>	<b>48</b>	<b>1 249</b>

**TABLE 5.19 – Results from Request 10**

The success rate of each search engine was now determined by adding the total number of bold figures (i.e. successful queries) from each of the columns of the above 10 tables, yielding the table below:

	AV	IS	YA	LY	HB	WC	EX
<b># OF QUERIES</b>	17	15	11	4	23	12	15

**TABLE 5.20 – Search Engine Success**

The two least successful search engines (Yahoo! and Lycos) were subsequently dropped from the last stage of experimentation.

The percentage success for each of the two categories of operators identified above (Inclusion/Phrases and Logical/Field) was now calculated.

The total number of bold figures (i.e. successful queries) from each of the rows of the above ten tables was expressed as a percentage of the number of queries, yielding the table below:

OPERATOR TYPE	SUCCESS/140	SUCCESS %
INCLUSION/PHRASE	70	50.0
LOGICAL/FIELD	27	19.3

**TABLE 5.21 – Operator Category Success**

It was clear from Table 5.21 that inclusion and phrase operators produced more relevant answers than the use of logical and field operators.

The final set of the author's searching experiments aimed at the design of FOIOTI will focus on the use of only the five selected search engines (see Table 5.20 and comment) and the selected search operators (see Table 5.21 and comment). The next ten REQUESTS were taken verbatim from ten of the forms completed by ten of the participants from the study. The queries were constructed by the author using the operators as discussed, and tested only on the remaining five search engines.

**REQUEST 11** (Participant number 153)

I need a copy of Visual Basic 5.0 code that enables motion in graphics.

**QUERY 11.1**

+ "Visual Basic 5.0" +source +code +graphics +motion

	AV	IS	HB	WC	EX
Q11.1	0	0	2 260	0	13

**TABLE 5.22 – Results from Request 11**

**REQUEST 12 (Participant number 162)**

I need to know how to write a Pascal program.

**QUERY 12.1**

+ "Pascal" +basic +tutorial

	AV	IS	HB	WC	EX
Q12.1	87	10	8 180	72	2 756

**TABLE 5.23 – Results from Request 12**

**REQUEST 13 (Participant number 193)**

C++ source code on arrays.

**QUERY 13.1**

+ "C++" + "source code" +arrays +tutorial

	AV	IS	HB	WC	EX
Q13.1	6	0	4 760	15	861

**TABLE 5.24 – Results from Request 13**

**REQUEST 14 (Participant number 207)**

I need the complete interrupt function list for programming the Microsoft Mouse in DOS using C++.

**QUERY 14.1**

+ "C++" + "Microsoft Mouse" +function +list +reference

	AV	IS	HB	WC	EX
Q14.1	0	0	92	1	44

**TABLE 5.25 – Results from Request 14**

**REQUEST 15 (Participant number 242)**

I need information of [sic] case studies on MIS.

**QUERY 15.1**

+ "MIS" +case +study +studies

	AV	IS	HB	WC	EX
Q15.1	0	14	3 550	71	3 430

**TABLE 5.26 – Results from Request 15**

**REQUEST 16 (Participant number 287)**

How to display a bitmap in a window using Visual C++.

**QUERY 16.1**

+ "Visual C++" +bitmap +display +window

	AV	IS	HB	WC	EX
Q16.1	0	0	5 160	9	296

**TABLE 5.27 – Results from Request 16**

**REQUEST 17 (Participant number 82)**

Database. I need to find out how SQL transmission management statements work and how are they implemented.

**QUERY 17.1**

+ "SQL" +transmission +management +statement +tutorial

	AV	IS	HB	WC	EX
Q17.1	14 210	215	310	5	0

**TABLE 5.28 – Results from Request 17**

**REQUEST 18 (Participant number 117)**

I need to know how to insert a PowerPoint animation into a webpage.

**QUERY 18.1**

+ "PowerPoint" + animation + insert + Web

	AV	IS	HB	WC	EX
Q18.1	0	0	250	2	398

**TABLE 5.29 – Results from Request 18**

**REQUEST 19 (Participant number 119)**

How exactly does Dijkstra's algorithm work to create a routing table?

**QUERY 19.1**

+ "Dijkstra" + "routing table" + algorithm

	AV	IS	HB	WC	EX
Q19.1	221	77	270	2	141

**TABLE 5.30 – Results from Request 19**

**REQUEST 20 (Participant number 86)**

How to draw and use Gant [sic] charts.

**QUERY 20.1**

+ "Gantt chart" + tutorial

	AV	IS	HB	WC	EX
Q20.1	274	34	88	1	63

**TABLE 5.31 – Results from Request 20**

#### 5.7.2.4 SEARCH ENGINE TEST RESULTS

The successes (bold figures) per search engine of the previous ten tables were added, producing the following final number of successes:

	AV	IS	HB	WC	EX
<b># OF SUCCESSES</b>	<b>3</b>	<b>3</b>	<b>7</b>	<b>4</b>	<b>7</b>

**TABLE 5.32 – Search Engine Success**

It was finally concluded that the operation of Model 3 would be based on the following:

- common errors identified in searching strategy (see the three types of errors, Section 5.3.2)
- the most successful operators, i.e. Inclusion and Phrases (Table 5.21), and
- the three most successful search engines, when used in combination with these operators, i.e. Hotbot, Webcrawler and Excite (Table 5.32). FOIOTI was coded to request five components from the user, which were combined with these operators and submitted to these search engines.

### 5.8 MODEL 3 (FOIOTI)

#### 5.8.1 DESIGN CRITERIA

During the reading phase of this project a number of decisions were taken regarding FOIOTI's design. A summary to confirm the implementation of these decisions is given below, in conjunction with the points listed in Section 5.7.2.4 above.

### 5.8.1.1 BASIC APPROACH

Forcese and Richer defined a model as "... an imitation or an abstraction from reality that is intended to order and to simplify our view of that reality while still capturing its essential characteristics" (Forcese and Richer 1973:38).

FOIOTI appears to be a search engine ("... imitation ... of reality"), it hides the technicalities of search engines ("... simplify our view ..."), but still provides useful answers to users' queries ("... still capturing its essential characteristics").

Furthermore, Fidel's identification of the two basic approaches to searching guided the way FOIOTI operates. While the operationalist approach stresses the "technology" of the search engine, including searching methods, the conceptualist approach changes the original search specification in an attempt to think around the concept of what the search topic actually is (Fidel 1991c:517). FOIOTI draws the focus of the searcher away from the searching technology, search engine syntax and operators, and frees the user to concentrate on the topic. If the user does not find enough information in his/her original specification to fill in FOIOTI's boxes, he/she simply changes or expands the original specification, similar to the method the conceptualist approach prescribes.

Finally, a number of authors proved that, for example, Boolean operators are either not used, or often used incorrectly by information seekers (Siegfried, Bates and Wilde 1993:273). Hirsh found that children experience problems when working with Boolean operators (Hirsh 1999:1266,1271). Various other problems were found with operators, such as the fact that some functions and operational logic sometimes produced inconsistent answers (Snyder and Rosenbaum 1999:380). FOIOTI hides these technical issues from the user.

### 5.8.1.2 EXPERIMENTAL METHODOLOGY

Two of the most fundamental works in the field of IR, namely the Cranfield tests and Lancaster's work on MEDLARS, relied on an experimental methodology similar to the one used for this study.

During the Cranfield tests, for example, extensive searching was done on a defined set of documents, and decisions were taken on the relevancy of the results (Cleverdon, Mills and Keen 1966, Robertson 2001, Tonta 2001).

FOIOTI's design was based on the results of a set of extensive searching experiments done on a set of documents (the Internet), and the searchers took decisions on the relevancy of the results. Furthermore, FOIOTI is also based on Fidel's statement that "... analyzing users' seeking and searching behavior as it occurs in actual situations is a promising method for ... suggesting improvements in system design and in search environments" (Fidel and others 1999:36).

#### 5.8.1.3 QUERY FORMULATION

Participants in earlier searching experiments could not be convinced to engage in the complex cognitive strategies required for IR (Wallace, Kupperman and Krajcik 2000:100). FOIOTI removes the need for the user to become involved in these strategies, since it generates a query based on the user's loose searching specification. Other authors have done a variety of successful tests on systems similar to FOIOTI. They attempted to apply an algorithm to formulate a query on the user's behalf, and in the process isolate the user from the technical issues involved in IR (Chen and others 1998:582,600,601, Gauch and Smith 1993:124-136). Taylor's identification of four different levels of question formation (visceral, conscious, formalised and compromised need) also played a part in FOIOTI's design (Ingwersen 1992:112,113). FOIOTI does the conversion of the user's formalised need (searching specification) into the compromised need (searching query) which is coded in the language of the searching system.

#### 5.8.1.4 TRAINING SESSION

Various authors emphasised the importance of a training session on the use of search tools, even though it is short. Hersh, Pentecost and Hickam proved that such a session levelled the playing field when comparing two diverse searching methods such as Boolean operators and natural language searching (Hersh, Pentecost and Hickam 1996).

Large and Beheshti's experiments with 50 schoolchildren proved that they exhibited sophistication in their IR strategies, admittedly after having been shown at least two basic search methods (Large and Beheshti 2000:1076).

Another study on a similar audience proved that a short training session minimised the negative effect of searching technicalities (Wallace, Kupperman and Krajcik 2000:75). The author decided to do a short training session before expecting the participants to use FOIOTI.

#### 5.8.1.5 MULTIPLE INDICES

A number of authors suggested that more than one index should be used during Internet searching to ensure reliable results (Chun 1999:137, Garman 1999:75, Jacso 1999a:227, Jacso 1999b:99, Mickey 1999:79, Spink 1996:603,606, Williams 1996:246). FOIOTI gives the user the option of using any one or more than one of three search engines, thereby adding to the richness of the document set from which answers are to be extracted.

The Java source code of FOIOTI is given in Appendix L.

#### 5.8.2 DESCRIPTION OF OPERATION

A brief discussion of the function and features of FOIOTI follows (a screendump of FOIOTI is found in Appendix H, and it can be accessed online at <http://www.mwe.co.za>).

- The minimum amount of instructions are given – for more detailed help the user is referred to a Help page, where clear instructions and examples are listed.
- The user is asked to enter his/her exact, full specification in the "Spec Box". This specification had to be related to the user's field of academic study, and a number of examples from the IT/IS field were given. This requirement was included after the success of Oliver and Oliver, who proved that learning takes place when information seeking is based on specific needs.

They also proved conclusively that a higher degree of success can be expected during information seeking inside a given context, than without it (Oliver and Oliver 1997:520,525). The FOIOTI specification is analogous to the booklet on Global Warming used by Oliver and Oliver in their experiment. Although not mentioned to the user, this sentence is not used anywhere in the query FOIOTI eventually generates - it simply serves to force the user to think about what he/she is actually searching for. It also makes it easier for the user to find the detail required in the next four boxes.

- The user then has to type in any word, which is normally spelt with one or more capitals, into the "Caps box". This word will eventually be surrounded with quotes to preserve the capitalization, and preceded by an inclusion operator. For example, if the user types in the word Turkey (referring to the country), FOIOTI will pass it to the search engine as +"Turkey", to prevent references to the bird by the same name (normally written in lowercase form) from being included in the answers. The purpose of this word is to exclude the same word, which has a different meaning when typed with lowercase, and give more focussed answers. Examples of how this feature could distinguish between words follow: Windows (operating system) versus windows (glass panes), Apple (computer) versus apple (fruit), Stone (surname) versus stone (rock). Since the word must contain one or more capital letters, it is likely that the user will enter a specific word with high filtering capacity, for example FBI, Mercedes, UK, DVD, etc. It therefore prevents the use of common words (computer, cars, programming, news, etc). However, if a word without any uppercase letters is entered, it will not impede the technical functioning of FOIOTI in any way.
- In the "Phrase box", the user has to type in two or more consecutive, meaningful words from the Spec box. The sequence of these words is conserved by quotes, and the inclusion operator precedes the phrase. Once again the purpose is to exclude rather than include answers, by including at least one two-word phrase.

In the Help menus the user is advised to adjust the length of the phrase, depending on the number of answers received – a longer phrase (i.e. more specific) gives fewer answers and vice versa.

- Finally, the user must type two other, non stop-words into the two “Word boxes”. These two words are left unquoted, to allow lower- and uppercase forms, and are also preceded by the inclusion operator. The purpose of these words is to allow the user to “fine tune” his query even more, although conversion to uppercase is enabled.
- From the experimental results above (Table 5.32) it is clear that Webcrawler, Hotbot and Excite responded best to the use of phrases included in quotes, and the use of the inclusion operator. As a result, they are listed as the search engines of choice in FOIOTI.

### 5.8.3 EXAMPLE

An example of how FOIOTI would respond to a user's input follows:

<b>Spec box:</b>	<b>I need to find the final death toll figure as a result of the earthquakes in Turkey during 1999.</b>
<b>Caps box:</b>	Turkey
<b>Phrase box:</b>	death toll
<b>Word box 1:</b>	earthquake
<b>Word box 2:</b>	1999
<b>FOIOTI's query:</b>	<b>+“Turkey” +“death toll” +earthquake +1999</b>

**TABLE 5.33 – Example of FOIOTI's Operation**

The three types of errors made by participants in the study up to this point identified in Section 5.3.2 were considered, and the model was designed to prevent all three. The Caps box does ask for a single word (one of the common mistakes identified earlier), but the correct use of one or more capital letters will filter out unwanted answers rather than add more superfluous ones.

When the user fills in the Phrase box, he/she should choose a sensible phrase from the Spec box, alleviating the problems (length, quotes, etc) identified earlier with the use of phrases.

Finally, since FOIOTI is syntax-free, users do not have to know even the generic syntax of search engines, since FOIOTI adds that automatically.

### **5.9            PHASE 11 – TEST FORM 4 WITH MODEL 3**

Another countrywide tour was undertaken to implement Model 3. The same three-component experiment of Phase 7 was executed, now using Form 4 with Model 3. To enable the users to record detail about their searching with FOIOTI, Form 5 was designed to make this possible (see Appendix E). Note that the initial URL listed in Box A of Form 5, has since been replaced by [www.mwe.co.za](http://www.mwe.co.za).

This phase involved 272 learners over 11 sessions at ten South African institutions, spread over six cities.

SESSION NUMBER	CITY	INSTITUTION	NO. OF LEARNERS
28	Bloemfontein	University of the Free State	10
29	Bloemfontein	Technikon Free State	18
30	Bloemfontein	Vista University – Bloemfontein	48
31	Welkom	Vista University – Welkom	48
32	Johannesburg	Vista University – East Rand	16
33	Johannesburg	Vista University – Soweto	32
34	Durban	University of Durban- Westville	22
35	East London	Border Technikon	31
36	Port Elizabeth	PE Technikon	27
37	Port Elizabeth	PE Technikon	34
38	Port Elizabeth	Vista University – PE	34
<b>TOTAL</b>	<b>11</b>	<b>6</b>	<b>10</b>
			<b>272</b>

**TABLE 5.34 – Phase 11 Session 28 – 38**

#### 5.10 PHASE 12 – INTERPRET RESULTS, DETERMINE SUCCESS OF MODEL 3

The most basic purpose of this research project was to enable a learner to find relevant information on the Internet. One way to measure FOIOTI's success in achieving this goal, is to compare the number of Yes answers before FOIOTI with the number of Yes answers thereafter.

This led to the definition of four possible classes, depending on the final answer given for each experiment (before/after): No/No, No/Yes, Yes/No and Yes/Yes.

However, since the two related experiments were often done on separate days at the same institution, it could happen that a learner attended only one of the two sessions. If one of the two answers were absent, that participant's contribution would become invalid, producing a fifth class. Other reasons for classifying a response as invalid included incomplete forms and invalid Yes/No answers, pointed out by other information supplied on the same form.

These five classes were now formalised. For ease of interpretation, they were arranged from most wanted result (Class 1) to least wanted result (Class 4).

CLASS	ANSWER (BEFORE/AFTER)	IMPLICATION
1	No / Yes	FOIOTI had a positive effect on searching ability
2	Yes / Yes	FOIOTI could not have a positive effect on searching ability – learner found information without it
3	No / No	FOIOTI had no positive effect on searching ability
4	Yes / No	FOIOTI had a negative effect on searching ability
5	Not applicable	Data was ignored

**TABLE 5.35 – FOIOTI Class Definition**

Finally, if class 5 was indicated, a reason was added.

NO.	DESCRIPTION
R1	Documentation incomplete – only did one experiment
R2	Information on Form incomplete – cannot decide if answer is Yes or No
R3	Information on Form strongly suggests that wrong Yes/No answer was given
R4	Other reasons

**TABLE 5.36 – FOIOTI Class 5 Reasons**

The results produced by this Phase are now classified according to these classes. Overall results:

FOIOTI CLASS	NUMBER	PERCENTAGE
1	76	38.8
5	63	32.1
3	31	15.8
2	23	11.8
4	3	1.5
<b>TOTAL</b>	<b>196</b>	<b>100</b>

**TABLE 5.37 – Phase 12 – FOIOTI Class results (5x)**

However, only class 1, 3 and 4 provide a true indication of FOIOTI's performance. In the case of class 2, FOIOTI could not have any positive effect on the learner's searching ability, and it did not have a negative effect. Class 5 is invalid for reasons outside the control of the author, and it has no bearing on FOIOTI's performance. Therefore both classes 2 and 5 are omitted for the final evaluation. Some of these invalid cases are discussed in detail in Chapter 6 – please see Section 6.5.2.

Performance results:

FOIOTI CLASS	NUMBER	PERCENTAGE
1	76	69.1
3	31	28.2
4	3	2.7
<b>TOTAL</b>	<b>110</b>	<b>100</b>

**TABLE 5.38 – Phase 12 – FOIOTI Class results (3x)**

It can thus be concluded that, for this phase, FOIOTI improved the searching ability of 69.1% of learners, it had no positive effect on 28.2% of learners, and it had a negative effect on 2.7% of learners.

#### 5.11 PHASE 13 – IMPLEMENT MODEL 3 ON CONTINUOUS BASIS

To conclude this series of experiments, it was considered necessary to implement FOIOTI with learners at the Cape Technikon, to confirm that it has value on an on-going basis. The three-component experiment of Phase 11 was repeated with standard classes, consisting of learners in their first and second years of study from the National Diploma in Information Technology and National Diploma in Financial Information Systems courses. All other control elements were applied as before, including 30 minutes allowed for each searching exercise, a 90-minute training session, etc.

SESSION NUMBER	CITY	INSTITUTION	NO. OF LEARNERS
39	Cape Town	Cape Technikon	24
40	Cape Town	Cape Technikon	26
41	Cape Town	Cape Technikon	36
42	Cape Town	Cape Technikon	49
43	Cape Town	Cape Technikon	65
44	Cape Town	Cape Technikon	43
45	Cape Town	Cape Technikon	30
46	Cape Town	Cape Technikon	9
<b>TOTAL</b>	<b>8</b>	<b>1</b>	<b>282</b>

**TABLE 5.39 – Phase 13 Session 39 – 46**

Overall results:

FOIOTI CLASS	NUMBER	PERCENTAGE
5	198	70.2
1	49	17.4
2	18	6.4
3	14	5.0
4	3	1.0
<b>TOTAL</b>	<b>282</b>	<b>100</b>

**TABLE 5.40 – FOIOTI Class results (5x)**

The high number of learners returning invalid responses (class 5) was a direct result of network denial-of-service problems experienced by the Cape Technikon at the time of those specific experiments. It made it impossible for a number of students to log onto the network, and as a result they could not use the Internet. If network problems or lack of Internet access prohibited the use of search engines, time did not allow the repetition of the session, and the forms were simply handed in without any searching having been done. This was the case with Session 42 and 43, already a total of 114 learners. Most of the other class 5 learners were those who attended only the first or only the second experiment.

After omitting the results for class 2 and 5, FOIOTI's performance-based results are:

FOIOTI CLASS	NUMBER	PERCENTAGE
1	49	74.2
3	14	21.2
4	3	4.6
<b>TOTAL</b>	<b>66</b>	<b>100</b>

**TABLE 5.41 – FOIOTI Class results (3x)**

To summarise FOIOTI's performance-based results of phase 11 and 13:

	PHASE 11 (PARTICIPANTS)	PHASE 11 (%)	PHASE 13 (PARTICIPANTS)	PHASE 13 (%)	AVE- RAGE
<b>CLASS 1</b>	76	69.1	49	74.2	<b>71.0</b>
<b>CLASS 3</b>	31	28.2	14	21.2	<b>25.6</b>
<b>CLASS 4</b>	3	2.7	3	4.6	<b>3.4</b>
<b>TOTAL</b>	<b>110</b>	<b>100</b>	<b>66</b>	<b>100</b>	<b>100</b>

**TABLE 5.42 – FOIOTI Class results (3x)**

Overall, FOIOTI thus improved the searching ability of 71.0% of learners, it had no positive effect on 25.6% of learners, and it had a negative effect on 3.4% of learners.

## **5.12 CLASS 4 RESULTS**

An inspection of the results of the six participants from Class 4 (the 3.4%) follows. These six all found information when searching without help, but did not find it while using FOIOTI.

### **5.12.1 RESULTS**

Participant number 517

Form 3 block F, G and H lead the participant through the identification, description and definition of a problem he/she was experiencing with academic work. The rest of the form then expected him/her to search for information, which would solve this problem. The form where users record their experience with FOIOTI also required that the search specification be noted down. In both cases, participant number 517 chose to search for non-academic material. When working with FOIOTI, he used only one search engine out of the three, visited seven websites and spent 22 minutes before giving up. One of his comments was: "It got little confusing [sic] working with FOIOTI".

Participant number 782

This participant used all three FOIOTI search engines, visited three websites and spent 20 minutes searching before giving up. No comments about his failure were given.

Participant number 783

During the first experiment, this participant specified "Development Software" as a topic, but actually appeared to search for employment in the IT industry. During the FOIOTI search, he did not specify his search topic, used only one search engine, visited only one site and gave up after only five minutes.

Participant number 855

This participant used a wide specification ("programming – C++"), claimed to have visited one site and found his information in two minutes during the first experiment. His FOIOTI search appeared to be quite exhaustive (only one search engine, but four websites in a claimed 40 minutes), but produced no results.

Participant number 864

This participant again used a very wide specification in the Mathematics subject ("Information on logs") and claimed to have found it in five minutes. During the FOIOTI search he again spent only five minutes before giving up, searching for a non-academic topic.

Participant number 1032

This user typed in a very wide topic as a single word: "COBOL", visited three sites and found information in five minutes. The FOIOTI search appeared to be exhaustive, involving all three search engines, lasting 25 minutes and covering ten sites.

### 5.12.2 ANALYSIS

A number of points from the results of the six users warrant discussion.

#### 5.12.2.1 WIDE SPECIFICATION

It is generally accepted that a claim of searching success with a very wide specification can be viewed with suspicion. For example, an imaginary user lists a search specification as "I need information on Java", and types in "Java" in the search box of a search engine. AltaVista produced 7 964 221 answers, including websites containing Java programming code, tutorials, lists of websites, and references to an island (AltaVista 2000). Since the user did not specify what exactly about Java needs to be determined, it is difficult to accept a claim of success in his/her searching bid. Participant numbers 783, 855, 864 and 1032 all used very wide specifications, and claimed success in their first searching experiment. This component of their response is therefore under suspicion.

#### 5.12.2.2 TIME USAGE

In all 46 sessions of searching experiments, the time limit for the actual searching component was personally timed by the author to be 30 minutes. In some cases users gave up well before this limit was reached during the second experiment, without supplying a reason. Five users did not use the full time available, while two of these only tried for five minutes each. For this reason, their "NO" answers in the FOIOTI experiment could be as a result of a lack of motivation to persevere with searching.

#### 5.12.2.3 MULTIPLE INDICES

Various authors claim that more than one index should be used while searching the world wide web (Garman 1999:75, Jacso 1999a:227, Jacso 1999b:99). It is for this reason that FOIOTI includes the option to use up to three search engines in parallel. Three of the six participants used only one search engine, even after clear instructions to the opposite effect were given.

FOIOTI's help menus clearly state a user should use more than one search engine (Weideman 2000a). This factor could be partially responsible for the failure to find information using FOIOTI.

#### 5.12.2.4 CONCLUSION

The results from these six users prove that some of them did not try their best while doing the experiments, and the author is not unduly concerned about the negative implication on the operation of FOIOTI.

# CHAPTER 6

## RESEARCH RESULTS AND ANALYSIS

### 6.1 INTRODUCTION

The purpose of this chapter was to summarise and analyse the results of the preceding seven sets of experiments. For clarity, the seven phases that involved empirical experiments with 1109 learners are tabulated below.

PHASE	DESCRIPTION	RAW RESULTS
1	Design and test Form 1	Appendix J: J.1
3	Test Form 2	Appendix J: J.2
5	Test Form 3	Appendix J: J.3
7	Test Form 3 and Model 1	Appendix J: J.4
9	Test Form 3 and Model 2	Appendix J: J.5
11	Test Form 4, 5 and Model 3	Appendix J: J.6
13	Implement Model 3	Appendix J: J.7

**TABLE 6.1 – Phase Summary**

It was decided not to distinguish between South African and overseas students, since that was not the original objective of the study. Furthermore, the sample of overseas students (90 learners) was small compared to the overall sample size (1109 learners).

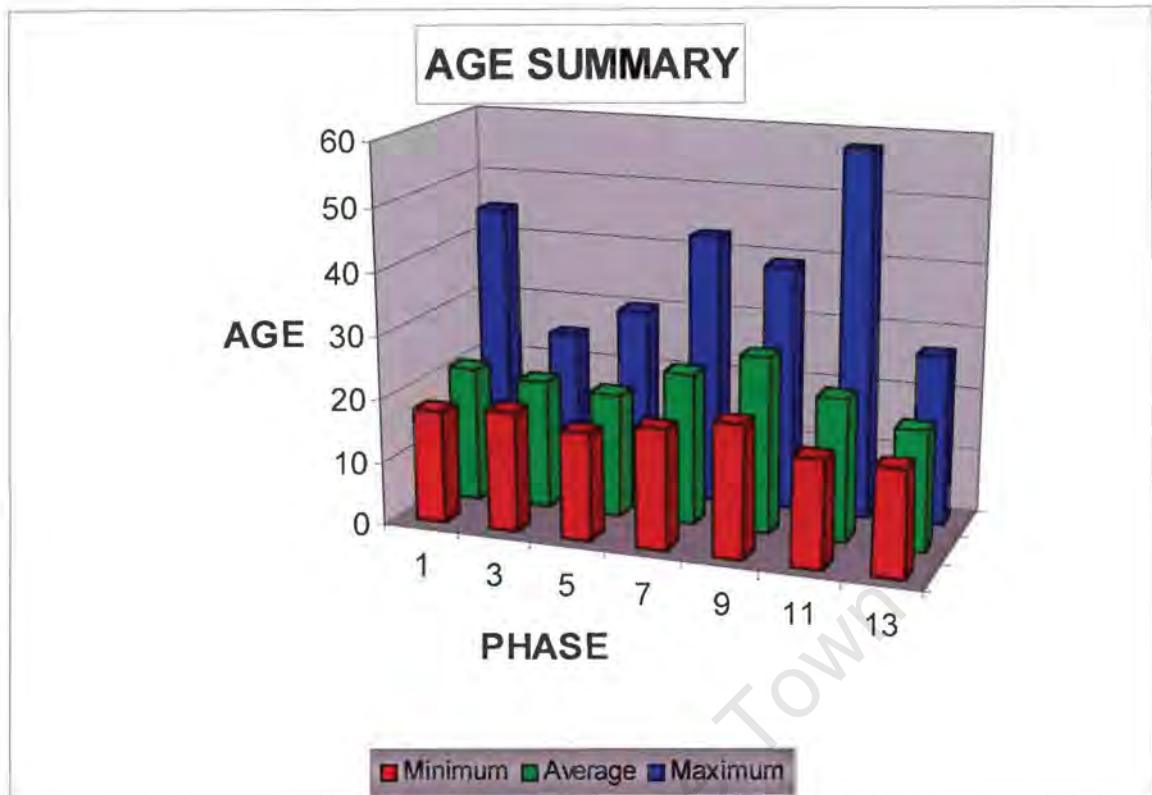
## 6.2 AGE

### 6.2.1 RESULTS

The age of each participant was requested on all three Forms, and is summarized below.

PHASE	MINIMUM	AVERAGE	MAXIMUM
1	18	21.5	45
3	19	20.6	25
5	17	19.7	30
7	19	24.1	43
9	21	27.8	39
11	17	22.6	58
13	17	19	27
<b>AVERAGE</b>	<b>18.3</b>	<b>22.2</b>	<b>38.1</b>

**TABLE 6.2 – Age Summary**



**GRAPH 6.1 – Age Summary**

### 6.2.2 ANALYSIS

The average age of 22.2 years was considered to be acceptable as being typical of learners having completed their school career, and busy with their first, second or third year of study. Statistical analysis has shown that the Coefficient of Variance (CV) for the seven experimental phases was: 6.8, 8.7, 11.1, 14.9, 17.2, 22.1 and 24.5 (see Appendix M, Section 6.2.2). Inspection of the two extreme values revealed that both those phases involved small sample sizes compared to the overall number of participants: Phase 3 (CV=6.8) had an  $n$  of only 17, while Phase 7 (CV=24.5) had an  $n$  of 55. Since the CVs of the different phases are relatively close, the samples are fairly homogeneous.

The average age of 22.2 years also compares favourably with what Corritore and others call “traditional students”.

The ages of a group of second-year computer science college students involved in a study done on laboratory-based approaches for teaching introductory computer science varied between 19 and 22 (Corritore and others 1999:295). In a study done on students' use of the Internet by Perry, Perry and Hosack-Curlin, it was found that regular use of the Internet was spread evenly over age groups from 18-19 years to 26 years (Perry, Perry and Hosack-Curlin 1998:140). Only 8.7% of the learners in the current study are outside the age group of 18 – 26 years of age.

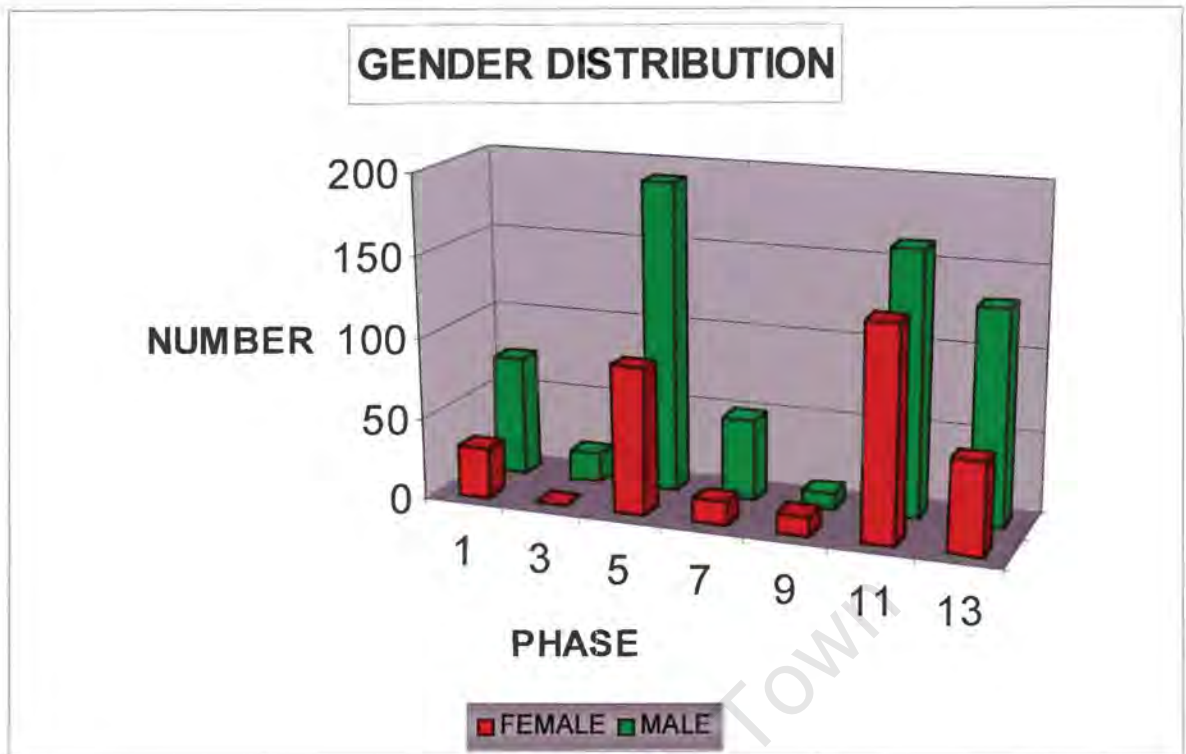
### 6.3 GENDER

#### 6.3.1 RESULTS

The gender of each participant was requested on all three Forms, and is listed below.

PHASE	FEMALE	MALE	UNKNOWN
1	31	74	0
3	1	18	0
5	91	191	11
7	15	50	3
9	12	10	0
11	131	162	27
13	56	131	95
<b>TOTAL</b>	<b>337</b>	<b>636</b>	<b>136</b>
<b>%</b>	<b>34.6</b>	<b>65.4</b>	<b>-</b>

**TABLE 6.3 – Gender Summary**



**GRAPH 6.2 – Gender Distribution**

### 6.3.2 ANALYSIS

The ratio of 1:1.89 (female:male = 34.6:65.4) was considered to be acceptable, since it is very close to the average (1:1.70) of three other sources: Applebee and others reported 1:2.30 in a study on Australian academics, Corritore and others noted a ratio of 1:1.5 in a study with college students and the Committee of Technikon Principals reported a figure of 1:1.30 for the 1998 academic year (Applebee and others 2000:143, Corritore and others 1999:295, CTP 1999:22).

## 6.4 RACE

### 6.4.1 RESULTS

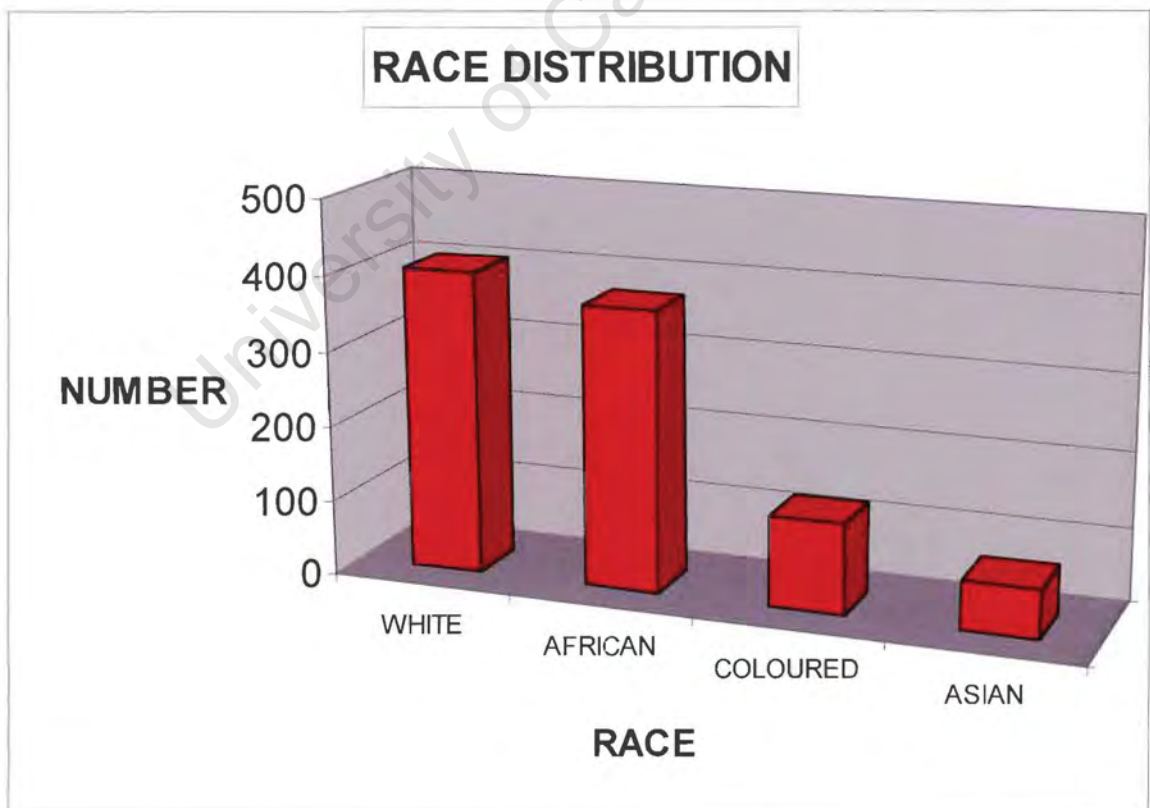
This study was not specifically aimed at South African learners, but by far the largest percentage of participants in this study were from South African institutions.

The changing nature of the South African education system prescribed that all four population groups identified by a South African Post-Secondary Education (SAPSE) document be represented in the sample (CTP 1999:34). This document presents a profile of all South African technikons, and is updated annually.

The race of participants was requested on all forms, although it is a sensitive topic in South Africa. It is also standard practice on most administrative forms of educational institutions. In their case it is done to determine whether equality policies are having the desired effect. As far as this study is concerned, the data was requested simply to ensure that the major population groups present in the country were all represented in the sample. Some students in the country are from disadvantaged backgrounds from the pre-1994 era (before the major election), and have obvious deficiencies in basic skills. Even those from post-1994 schooling systems are still suffering the consequences of educational deprivation of earlier years. Care was taken to include learners from this background in the sample.

PHASE	AFRICAN	ASIAN	COLOURED	WHITE	OTHER	UNKNOWN
1	5	5	35	60	0	0
3	0	0	1	18	0	0
5	138	30	8	110	0	7
7	1	0	2	63	0	2
9	1	6	0	11	11	1
11	215	19	10	43	1	32
13	13	5	66	98	2	90
<b>TOTAL</b>	<b>373</b>	<b>65</b>	<b>122</b>	<b>403</b>	<b>14</b>	<b>132</b>
<b>%</b>	<b>38.2</b>	<b>6.7</b>	<b>12.5</b>	<b>41.2</b>	<b>1.4</b>	<b>-</b>

**TABLE 6.4 – Race Summary**



**GRAPH 6.3 – Race Distribution**

#### 6.4.2 ANALYSIS

Both the "Other " and "Unknown" figures were omitted from the graph, since the "Other" figure is too small to be of significance, and the "Unknown" figure has no bearing on the distribution. The exact figures for race distribution of learners world wide and in South Africa are not known. However, the same four broad population groups for South Africa as in the graph above were listed in a current encyclopaedia, with the following values: African 75.2%, White 13.6%, Coloured 8.6% and Asian 2.6% (Microsoft 2000). The difference between the "White" and "African" values in these two sets of figures was a result of the difference in the population composition of the whole country, and the population composition of higher education institutions in the country. This study has no political or historical component, but this difference can be attributed to the history of discrimination in the South African education system.

However, the author considered the composition of the learners in this study, as depicted by the graph above, to be representative of IT/IS learners in the country. This was confirmed by the fact that a wide variety of higher education "types" were included in the study. Traditionally "white", advantaged Institutions were: Cape Technikon, Natal Technikon, Port Elizabeth Technikon, Technikon Free State, Rand Afrikaans University, University of the Free State, Stellenbosch University and University of Natal (Pietermaritzburg). Traditionally "black", disadvantaged institutions were Witwatersrand Technikon, Border Technikon, University of Durban-Westville and the following six Vista University campuses: Welkom, East Rand, Soweto, Mamelodi, Port Elizabeth and Bloemfontein. This classification of advantaged and disadvantaged institutions was based on the author's perceptions after having visited each one of the institutions – some were visited twice. At every institution the author actually worked with the learners, had intense discussions with the controllers, saw and used the facilities, and spent between three hours and two days on each campus. Furthermore, seven of the nine disadvantaged institutions mentioned above, were identified as having been founded with the sole purpose of providing higher education for racially disadvantaged groups (NCHE 2000). De Jager confirmed that the six Vista Universities were started specifically for black students (De Jager 1995:104)

At the same time, Robbins claimed that at least three of the eight advantaged institutions identified by the author above were so-called HAIs: Historically Advantaged Institutions (Robbins 1999). De Jager also listed another one of the HAIs identified by the author above (De Jager 1995:103). It was therefore concluded that the race distribution of the sample closely mirrors the real situation in South African institutions of higher education.

## 6.5 VALIDITY

### 6.5.1 RESULTS

The detail considered up to this point had been demographic, and all 1109 participants were included. A number of these learners, however, returned invalid responses with respect to the actual experimental data completed (or omitted) on their forms.

PHASE	VALID	INVALID
1	102	3
3	17	2
5	236	57
7	55	13
9	20	2
11	196	124
13	123	159
<b>TOTAL</b>	<b>749</b>	<b>360</b>
<b>%</b>	<b>67.5</b>	<b>32.5</b>

**TABLE 6.5 – Validity summary**

### 6.5.2 ANALYSIS

The reasons for disqualifying the invalid responses were varied. In many cases the home page of the learner's institution was used as search engine URL (for example, participant numbers 214 and 305).

It was assumed that this was set as the home page on the browser, and after loading the browser and being faced with this home page, the user did not know what to do from that stage on. In other cases a URL with no bearing to searching was used (participant numbers 336, 338, 343 and 353). It was assumed that the users knew this URL, and for lack of knowing how to search, they typed that in.

Some other reasons for disqualifying the invalid forms were:

- Participant number 73: learner claimed that two results were supplied by the search engine, but five sites were visited – therefore search engine results were not used as only measure of effectiveness of the search.
- Participant number 78: learner did not specify any topic or keywords.
- Participant number 98: learner answered Yes, but according to her description, the information found did not satisfy the query.
- Participant number 110: the learner claimed that 845 websites were returned, claimed to have found the answer, but did not list the actual number visited.
- Participant number 831: learner claimed to have found the answer, but did not visit any sites.
- Participant number 842: learner claimed to have found the answer, but did not use any search engine – simply typed in the URL of the answer.
- Participant number 854: learner claimed to have found the answer, but did not visit any sites.

From this point onwards, the forms of the 360 invalid participants were ignored, leaving 749 valid forms. This number was well above the calculated minimum sample size of 585 learners (see Section 4.1.5).

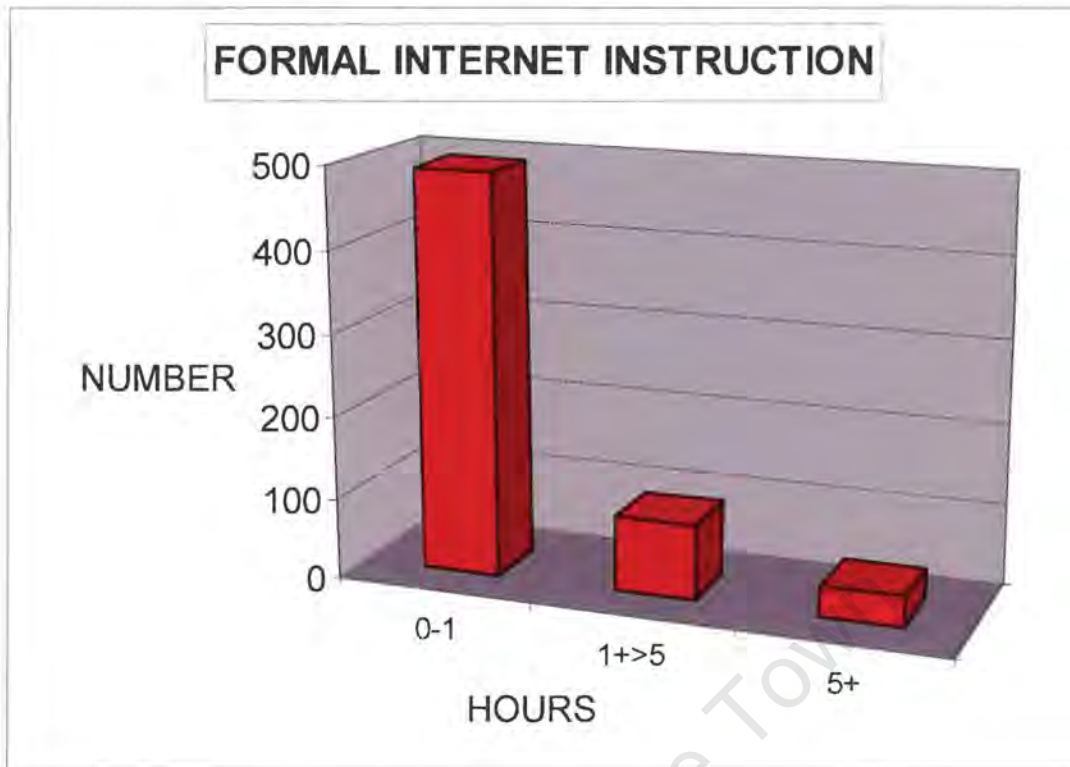
## 6.6 FORMAL INSTRUCTION

### 6.6.1 RESULTS

The question regarding the amount of formal instruction was asked to enable the author to determine the amount of formal Internet instruction received by the participants, and to be able to compare it with actual daily Internet usage (see Section 6.7). "Formal instruction" was defined as being instruction received on the use of the Internet at the institution where the experiments were done, under the guidance of a tutor or lecturer.

PHASE	1 (0-1 hours)	2 (1+ to 5 hours)	3 (more than 5 hours)	UNKNOWN
1	-	-	-	-
3	16	0	1	0
5	177	39	20	0
7	43	11	1	0
9	16	2	1	1
11	142	30	12	12
13	97	11	2	13
<b>TOTAL</b>	<b>491</b>	<b>93</b>	<b>37</b>	<b>26</b>
<b>%</b>	<b>79.0</b>	<b>15.0</b>	<b>6.0</b>	<b>-</b>

**TABLE 6.6 – Formal Internet Instruction Summary**



**GRAPH 6.4 –Formal Internet Instruction Summary**

### 6.6.2 ANALYSIS

The strong decrease in numbers from left to right in the graph indicated that most learners had no or little official training on Internet usage, while only a few had received official training. A total of 79% of participants claimed to have had one hour or less of formal instruction on the use of the Internet, while only 6% claimed having received instruction for more than five hours. It was therefore expected that the usage pattern of the next section would show the same trend – if eight out of ten learners received virtually no official training on the use of the Internet, those eight would probably not use the Internet regularly.

These results have a parallel in a similar study done at universities in the Netherlands, where Voorbij found that "... most Internet users learned to use the Internet services without the aid of a course;..." (Voorbij 1999:599). A total of 79% of the participants of the current study claimed one hour or less of formal Internet training.

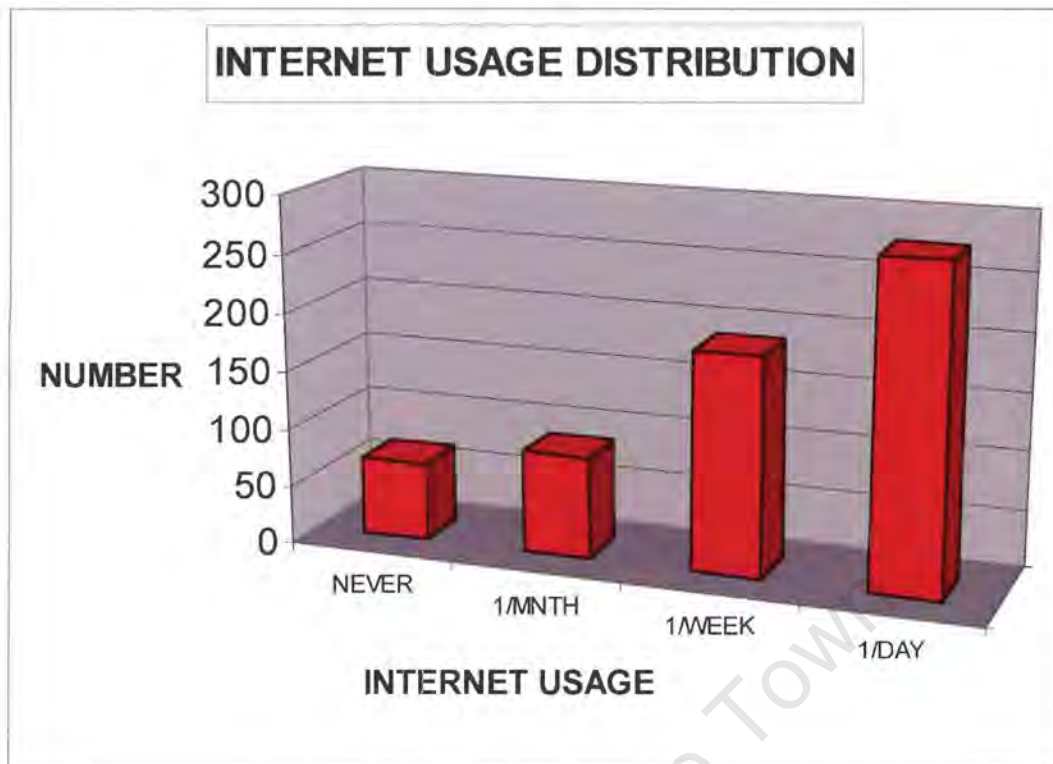
## 6.7 INTERNET USAGE

### 6.7.1 RESULTS

This question was asked to enable the author to determine the degree of exposure to Internet usage of the participants.

PHASE	1 (NEVER USED)	2 (APPROX. 1X/MONTH)	3 (APPROX. 1X/WEEK)	4 (AT LEAST 1X/ DAY)	UNKNOWN
1	-	-	-	-	-
3	2	1	4	10	0
5	10	25	81	120	0
7	4	11	15	25	0
9	0	0	3	17	0
11	33	22	44	85	12
13	19	30	42	20	12
<b>TOTAL</b>	<b>68</b>	<b>89</b>	<b>189</b>	<b>277</b>	<b>24</b>
<b>%</b>	<b>10.9</b>	<b>14.3</b>	<b>30.3</b>	<b>44.5</b>	<b>-</b>

**TABLE 6.7 – Internet Usage Summary**



**GRAPH 6.5 – Internet Usage Distribution**

### 6.7.2 ANALYSIS

The increase in numbers from left to right in the graph indicated that few learners (10.9%) have never used the Internet, increasing to the largest number (44.5%), who are using the Internet at least once per day. These figures were not in line with the expectation mentioned in Section 6.6.2 above – the two results in fact appeared to contradict each other. Formal Instruction results, in sequence, were: 1 2 3 (where 1 indicated little instruction), while Internet Usage results were: 4 3 2 1 (where 4 indicated high usage). Therefore a large group of respondents (79.0%) claimed to have had little or no official instruction on the use of the Internet, but at the same time 74.8% also claimed to use it at least once a week! This reverse association is significant at the statistical level of 0.1% (see Appendix M, Section 6.7.2). This phenomenon seemed to indicate that many Internet users from the learner group used the Internet regularly, without having been trained properly. This fact indicated that there was a possible need for the searching model – further evidence later would confirm this.

At this stage of each phase of the experiments, no actual searching results were available yet, which could possibly support the observation of the large, untrained group of Internet users. It was possible that the high percentage of users who were working on the Internet regularly, were doing so in an inefficient, unsatisfactory way. Furthermore, the claim of "using the Internet at least once per day" could, for example, simply imply the reading and sending of e-mail messages. Once again some correlation is evident with the Voorbij study: 80% of respondents claimed to have used the Internet (Voorbij 1999:599). The equivalent figure for this study is 89.1% (100% – 10.9%).

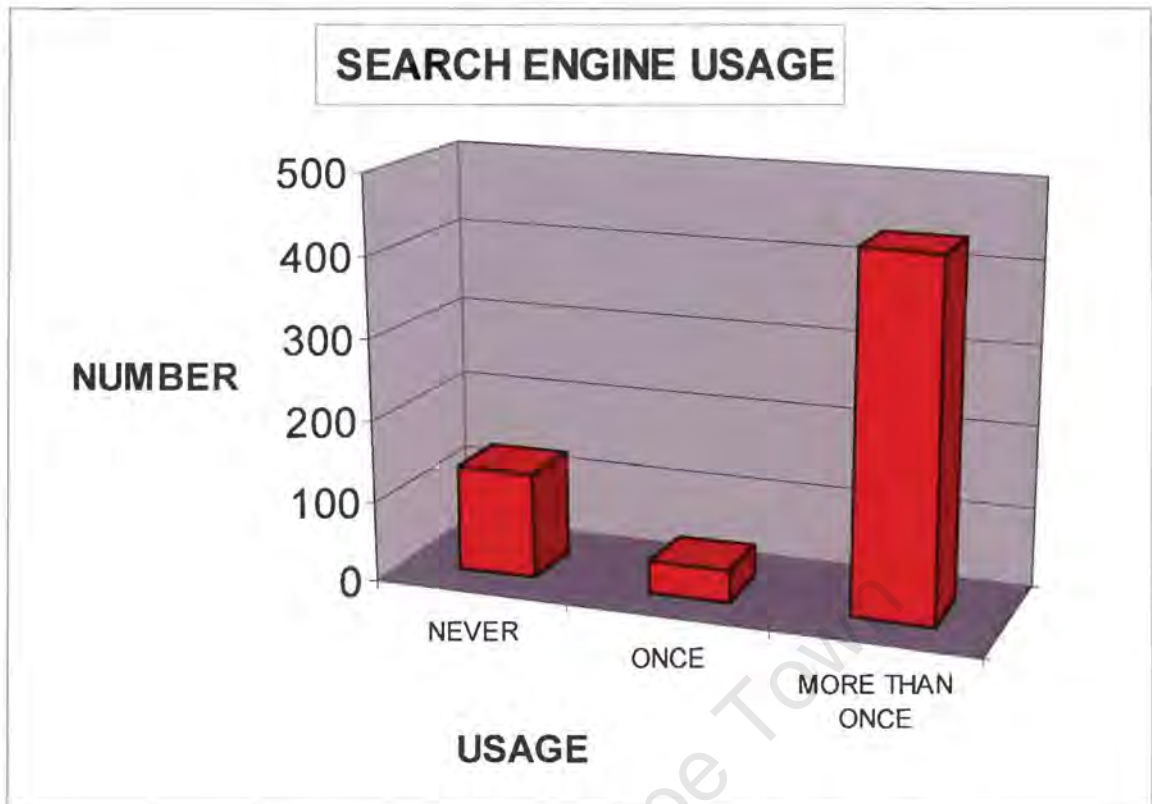
## 6.8 SEARCH ENGINE USAGE

### 6.8.1 RESULTS

The question about search engine usage was asked to enable the author to determine the amount of exposure to the use of Internet search engines the participants have had, and to be able to compare it with success achieved during searching.

PHASE	1 (never used)	2 (used once)	3 (used more than once)	UNKNOWN
1	-	-	-	-
3	-	-	-	-
5	30	13	193	0
7	4	2	49	0
9	0	0	19	1
11	69	15	102	10
13	28	9	74	12
<b>TOTAL</b>	131	39	437	23
<b>%</b>	<b>21.6</b>	<b>6.4</b>	<b>72.0</b>	-

**TABLE 6.8 –Search Engine Usage Summary**



**GRAPH 6.6 – Search Engine Usage summary**

### 6.8.2 ANALYSIS

A large percentage of learners (72%) claimed that they had used a search engine more than once before, while 21.6% had never used one. This seemed to suggest that seven out of 10 learners were at least familiar with the concept of searching on the Internet, and should know how and where to start a search.

If the “used once” and “used more than once” figures are combined, a total of 78.4% of participants use the searching facilities of the WWW at least occasionally. This corresponds with Voorbij, who claims a figure of 85% for the same category (Voorbij 1999:601).

## 6.9 SEARCH ENGINES

### 6.9.1 RESULTS

The question on search engine choice was included to determine which search engines learners most commonly used when given a free choice.

The data would also be used to make informed choices of which search engines to use in the model(s).

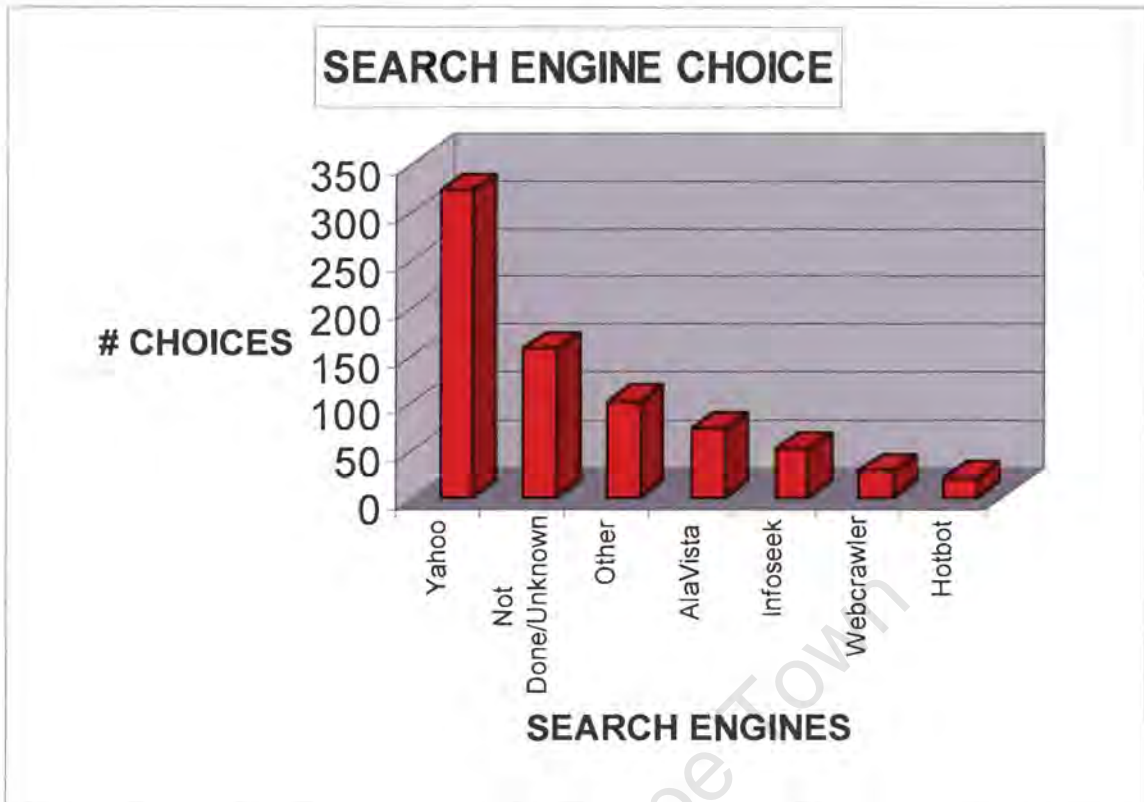
PHASE	37.COM	AARDVARK	ALL THE WEB	ANANZI	ASK JEEVES	ALTA VISTA	CYBER4
1	-	-	-	-	-	4	-
3	-	-	-	-	-	2	-
5	-	1	-	10	-	23	-
7	-	-	-	-	-	17	-
9	-	-	-	-	1	6	-
11	3	-	1	1	-	12	1
13	-	-	-	1	-	8	-
<b>TOTAL</b>	<b>3</b>	<b>1</b>	<b>1</b>	<b>12</b>	<b>1</b>	<b>72</b>	<b>1</b>
<b>%</b>	<b>0.51</b>	<b>0.17</b>	<b>0.17</b>	<b>2.02</b>	<b>0.17</b>	<b>12.14</b>	<b>0.17</b>

PHASE	EXCITE	GOOGLE	GOTO	HIGHWAY	HOTBOT	IBM	INFOSEEK
				61			
1	-	-	1	-	2	-	13
3	1	-	-	-	4	-	1
5	2	-	1	-	4	1	22
7	-	2	-	-	4	-	-
9	-	-	-	-	-	-	2
11	6	-	1	1	3	-	9
13	2	3	-	-	3	-	4
<b>TOTAL</b>	<b>11</b>	<b>5</b>	<b>3</b>	<b>1</b>	<b>20</b>	<b>1</b>	<b>51</b>
<b>%</b>	<b>1.85</b>	<b>0.84</b>	<b>0.51</b>	<b>0.17</b>	<b>3.37</b>	<b>0.17</b>	<b>8.60</b>

PHASE	LOOK SMART	LYCOS	MAMMA	META CRAWLER	MICRO SOFT	NETSCAPE SEARCH	NOT DONE
1	-	1	-	5	-	2	1
3	-	-	-	-	-	-	-
5	-	3	1	5	2	8	9
7	-	2	-	1	1	-	-
9	-	1	-	2	-	1	-
11	2	3	-	-	-	1	-
13	-	4	-	-	-	1	-
<b>TOTAL</b>	<b>2</b>	<b>14</b>	<b>1</b>	<b>13</b>	<b>3</b>	<b>13</b>	<b>10</b>
<b>%</b>	<b>0.34</b>	<b>2.36</b>	<b>0.17</b>	<b>2.19</b>	<b>0.51</b>	<b>2.19</b>	<b>-</b>

PHASE	SEARCH	SNAP	UN KNOWN	WEB CRAWLER	WEB FERRET	WEBO PEDIA	YA HOO	ZA ZOO
1	1	-	3	3	-	-	66	-
3	-	-	1	1	-	-	7	-
5	5	-	1	13	1	1	120	3
7	-	-	-	-	-	-	28	-
9	-	1	-	-	-	-	6	-
11	-	-	88	9	-	-	55	-
13	2	-	53	2	-	-	40	-
<b>TOTAL</b>	<b>8</b>	<b>1</b>	<b>146</b>	<b>28</b>	<b>1</b>	<b>1</b>	<b>322</b>	<b>3</b>
<b>%</b>	<b>1.35</b>	<b>0.17</b>	<b>-</b>	<b>4.72</b>	<b>0.17</b>	<b>0.17</b>	<b>54.29</b>	<b>0.51</b>

<b>TABLE 6.9 – Search Engine Summary</b>
--



**GRAPH 6.7 – Search Engine Summary**

### 6.9.2 ANALYSIS

A total of 27 different search engines were chosen without any guidance given from the author. Over 54% of the participants chose Yahoo! – more than all the other search engines put together.

In Voorbij's study, Yahoo! was the second most common choice out of a list of 20 search engines, after AltaVista (Voorbij 1999:603). If the "Not done/Unknown" and "Other" categories of the current study were ignored, these same two search engines occupy the first two positions in reversed order.

Statistical testing has proved that the difference in the proportion of participants choosing Yahoo! versus those choosing other search engines is significant at the 0.05 level (see Appendix M, Section 6.9.2).

Cohen and Jacobson noticed this phenomenon among learners and ascribed it to the misconception that Yahoo! contains all information available on the Internet (Cohen and Jacobson 1999:1). The author considered the catchy name and youthful image conveyed by the name as being part of the reason, as well as the fact that Yahoo! is one of the oldest search engines available.

Voorbij referred to the fact that "... meta-search engines ... are hardly used" (Voorbij 1999:603). This was confirmed in the current study, since the most often chosen meta search engine is only number 13 out of 27 search engines.

The learners' choices of search engines were considered in the choice of search engine to be used in FOIOTI.

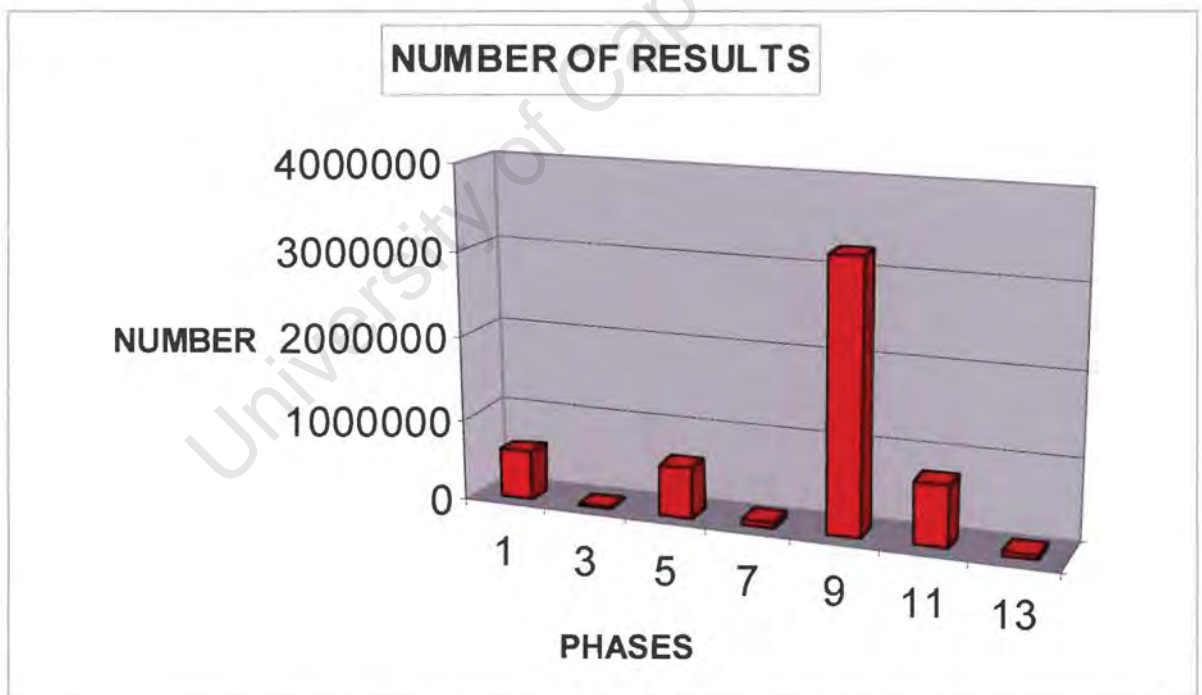
## **6.10        RESULTS RECEIVED**

### **6.10.1      RESULTS**

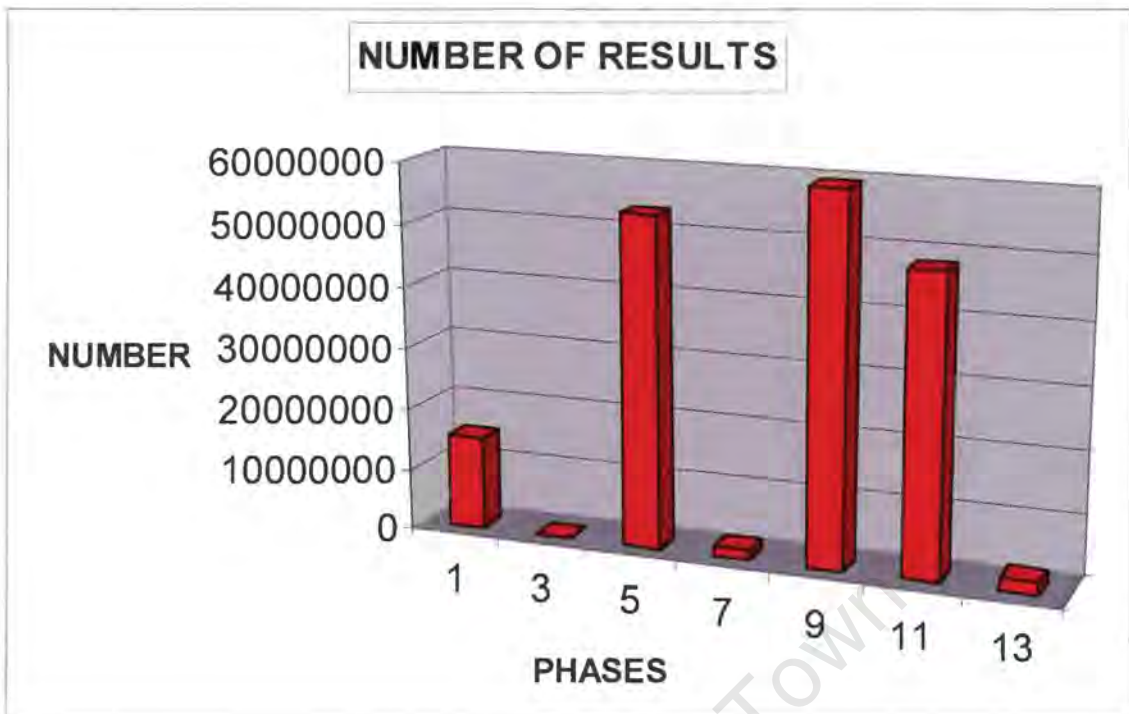
This question was included to determine the number of website summaries that was received, and to observe any changes in this figure as the testing of the model progressed. Lower numbers (between one and a few hundred) here would indicate higher competency in searching, since an exceptionally high number of answers in Internet searching (thousands and more) often indicated searching incompetence. A high number of answers can be obtained if a single, general word is used as query (for example: USA, computers, history, etc). Furthermore, incorrect syntax could also generate a high number of answers (for example: typing a sentence without quotes where the search engine does not support natural language searching, and adds an implied OR between words).

PHASE	MAXIMUM	AVERAGE	MINIMUM
1	14955214	584233	0
3	132542	7852	1
5	53469078	606348	0
7	1589213	72295	1
9	59510332	3241376	1
11	48337250	751831	0
13	2107225	58539	1
<b>AVERAGE</b>	<b>25728693</b>	<b>760353</b>	<b>0.57</b>

**TABLE 6.10 – Results Received Summary**



**GRAPH 6.8 – Results Received Summary: Average**



**GRAPH 6.9 – Results Received Summary: Maximum**

### 6.10.2 ANALYSIS

Some participants did not know how or where to search, resulting in the figure of 0 websites found. This Minimum figure was omitted from the graph, since it conveys no real meaning. Both the Average and Maximum numbers of summaries were prohibitively high, which is an indication that the participants were not applying appropriate methods to filter the number of answers down to a manageable number.

However, the overall maximum of 59 510 332 websites found by participant number 492 (phase 9), using Yahoo!, was viewed with suspicion. The index size of the seven biggest search engine databases (Google, AltaVista, FAST, Northern Light, Excite, Inktomi, and Go) at the time of this specific experiment varied between approximately 15 million and 130 million pages. Since Yahoo! was not included in this list, it implies that Yahoo!'s index was smaller than 15 million pages at that time (Sullivan 2000). It would thus have been impossible to achieve 59 million hits using Yahoo! at the time.

A clear decrease was noted in both the Average and Maximum figures over the last three phases. This decrease indicated that as the model was being refined, the ability of the learners to minimize the number of sites received, increased.

Statistical testing has confirmed the following: the decrease in the means from phase 9 to phase 11 and from phase 11 to phase 13 was not significant. However, the decrease in means from phase 9 to phase 13 was significant at the 0.05 level of significance (see Appendix M, Section 6.10.2).

The author thus concluded that the application of the model has had a positive effect on the searching ability of learners.

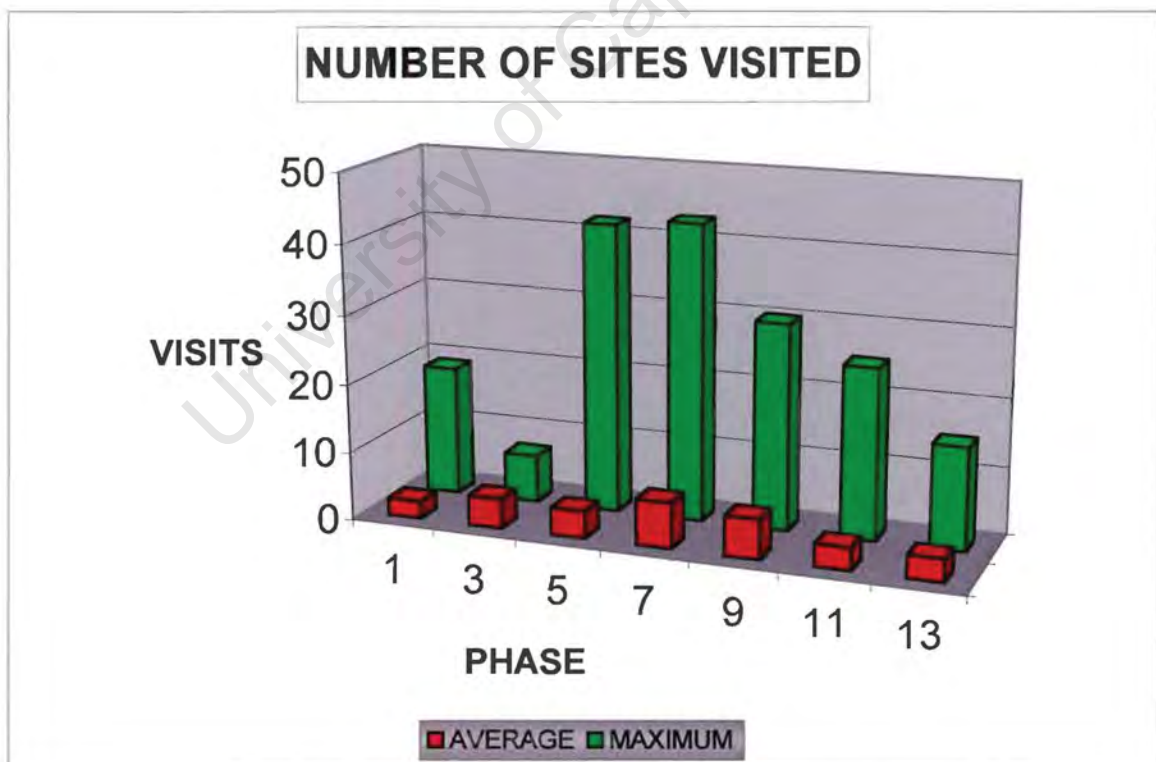
## **6.11        NUMBER OF WEBSITES VISITED**

### **6.11.1        RESULTS**

This question was included to determine the number of websites actually visited by the learner, and to observe any changes in this figure as learners became more efficient in their searching experiments. Lower numbers here would indicate higher competency in searching, since a consistently low number would indicate that a searcher focused his/her search narrowly by building up an exclusive query which filters out unwanted websites. This would increase the chance of finding the best answer after fewer visits.

PHASE	MAXIMUM	AVERAGE	MINIMUM
1	19	2.5	0
3	7	4.3	0
5	42	4	0
7	43	6.7	1
9	30	5.8	0
11	25	3.3	0
13	15	3.3	0
<b>TOTAL</b>	<b>181</b>	<b>29.9</b>	<b>1</b>
<b>AVERAGE</b>	<b>25.9</b>	<b>4.3</b>	<b>-</b>

**TABLE 6.11 – Number of Websites Visited Summary**



**GRAPH 6.10 – Number of Websites Visited Summary**

### **6.11.2 ANALYSIS**

Once again the 0 answer referred to the participants who did not actually execute any searches, and as before, this figure was omitted from the graph for clarity. The average of 4.3 sites for Phase 3 was higher than the 2.5 of Phase 1. This proved that the extra time freed by doing fewer searches could be used to search more extensively. The maximum having decreased from 19 to 7 indicated a higher success rate – it was not necessary to visit as many sites before success was achieved. Thus the problem of cramming three searches into one 30-minute session was solved (see Section 4.4.9).

A steady decline was noticeable in both the average and maximum from phase 7 to 13 on the graph. This indicated that as the model was being successively refined, learners needed fewer site visits to find the required information.

The author concluded that the model, even in its early versions, had a noticeable effect in increasing the chances of finding relevant information on the Internet by decreasing the number of websites visited.

## **6.12            NUMBER OF YES ANSWERS BEFORE MODEL**

### **6.12.1        RESULTS**

This was considered to be the most important feedback from Form 1, 2 and 3: how many of the participants actually found what they specified to be looking for within the time allowed?

PHASE	NO	YES	UNKNOWN	%YES
1	63	34	5	35.1
3	8	8	1	50.0
5	135	90	11	40.0
7	36	19	0	34.5
9	12	8	0	40.0
11	153	40	3	20.7
13	79	32	12	28.8
<b>TOTAL</b>	<b>486</b>	<b>231</b>	<b>32</b>	<b>-</b>
<b>%</b>	<b>67.8</b>	<b>32.2</b>	<b>-</b>	<b>32.2</b>

**TABLE 6.12 – Number of Yes Answers before Model Summary**

### 6.12.2 ANALYSIS

The average percentage of Yes answers before the exposure to any model (i.e. phase 1, 3 and 5) was 39.1% (132 Yes and 206 No answers). This meant that only four out of every ten learners at this stage claimed to have found one piece of academic information in 30 minutes, without any assistance.

This result seemed to be contradictory to the expectation mentioned in Section 6.8.2, where 72% of learners claimed that they had used an Internet search engine more than once before. However, it is in line with research literature – a brief summary of some of the findings of Chapter 2 follows.

A large number of authors claimed that they considered it important that learners should be able to find information on electronic sources:

- "...learning is more than ... simple retrieval of information" (Cronje and Clarke 1999:214).
- "... all students ... need to be able to locate, manage, evaluate and use information" (De Jager and Sayed 1998:197).

- "... students should know about the Internet ... as a tool and a source of information" (Edling 2000:11).
- "... to provide the opportunity for teachers and pupils to retrieve information from various sources around the world" (Gan 1998:29).
- "Speaker after speaker emphasised the need for the ... ability to find information required for solving problems" (Marais and Marais 1999:85-86).
- "They must learn to locate information in databases around the world, sieve through them, analyse what they get, and apply it to their projects in innovative ways. These will be key skills for every worker in future." Wong (1998:5) quoting Tong.

When comparing this set of statements with the next set, a strong contrast is noted. While the authors above consider it important that learners should be capable of IR on the Internet, the ones below state that IR for learners is fraught with difficulties.

- "Finding relevant and precise information on the Web can be extremely difficult" (Coetzee 1999:9).
- "... the results may well be disappointing, if not downright misleading" (Cohen and Jacobson 1999:1).
- "The very size of digital libraries begs the question of how average users will continue to find context and meaning" (Huwe 1999:68) .
- "The rich promise of Internet information is tantalizing, but usually frustrating" (Huwe 1999:68).
- "Would a savvy searcher know all these...? Hardly. A casual user such as a college student? No way" (Jacso 1999b:99).
- "Currently, search is simply bad" Sherman (1999:54) quoting Truher.
- "... some respondents seemed confused about what they were to report when asked to list query terms for their search" (Spink, Bateman and Jansen 1999:122).
- "...the user's ability to specify good search terms and create complex search queries to clearly and precisely capture relevant retrieval seems rather low" (Spink, Bateman and Jansen 1999:125).

- "...information seeking is a complex and difficult process ..." (Wallace, Kupperman and Krajcik 2000:75).

The original null hypothesis is as follows (from Section 4.1.2):

**H<sub>0</sub>: IT/IS learners can use the searching facilities on the Internet effectively without any aid to find reference materials, which will assist them in their studies.**

The author considered the 39.1% success rate mentioned above to be an unacceptably low figure, and concluded that IT/IS learners do not display enough insight to find relevant reference materials, unassisted, on the Internet. Consequently, this null hypothesis has been proven false.

The alternative hypothesis was postulated as (from Section 4.1.2):

**H<sub>1</sub>: IT/IS learners cannot use the searching facilities on the Internet effectively without any aid to find reference materials, which will assist them in their studies.**

One final calculation was made before proceeding with the proof of the alternative hypothesis. The 39.1% success rate was based on Yes/No answers of the first three phases only (Phases 1, 3 and 5), before the existence of any model. Since these Yes/No answers from the first experiment (searching without any training or guidance) were obtained before any exposure to any model in any case, the model, in whatever form, never affected a learner's Yes/No answer to this first search.

Therefore, as final confirmation of the lack of success in searching, the number of Yes/No answers of all seven sets of experiments was summarized, yielding an even lower figure of 32.2% of learners finding the information they specified (486 No, 231 Yes).

As a result of this low success rate, the author was convinced that it was necessary to proceed and design a method to assist learners in finding information on the Internet. Thus at this point  $H_1$  would have to be proved to be true, as discussed in Section 4.1.2.

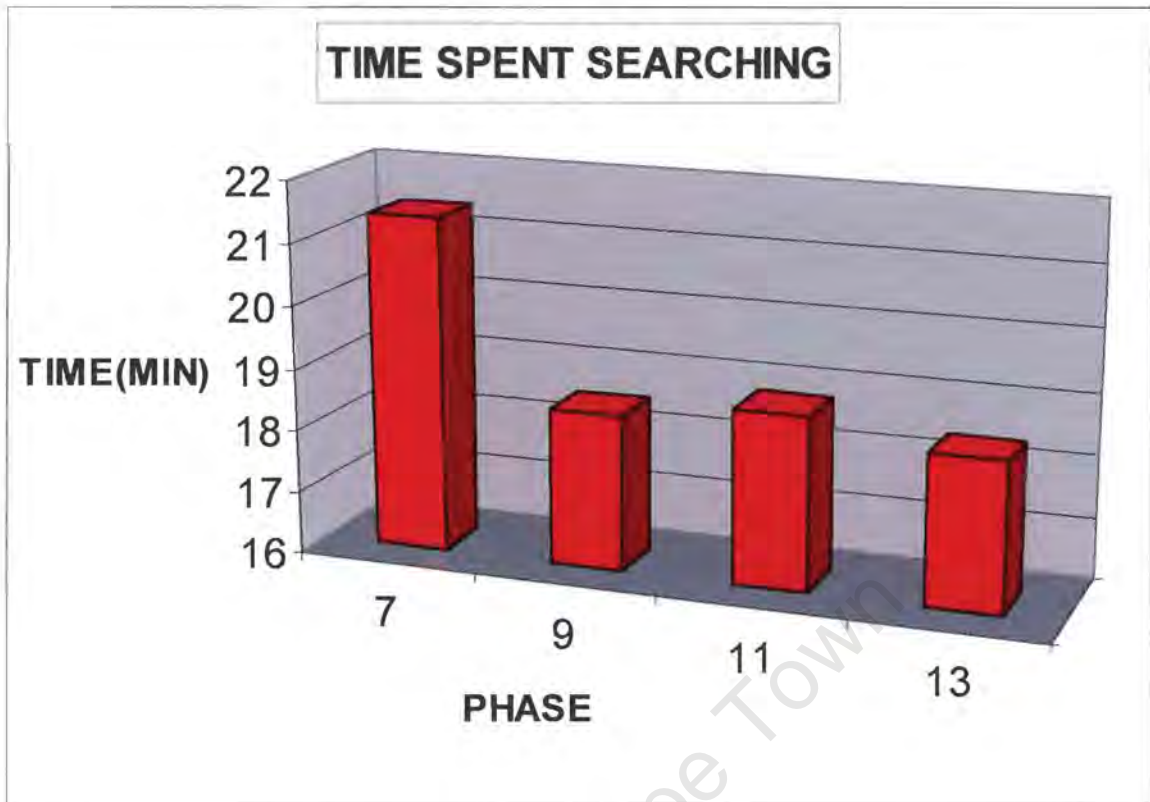
### 6.13 TIME SPENT SEARCHING

#### 6.13.1 RESULTS

The time (in minutes) each participant spent looking for the specified information was recorded here. Only the times of the participants who indicated success while using the model are shown in the table and the graph.

PHASE	AVERAGE TIME (minutes)
7	21.4
9	18.5
11	18.8
13	18.4
<b>AVERAGE</b>	<b>19.3</b>

**TABLE 6.13 – Time Spent Searching Summary**



**GRAPH 6.11 – Time spent Searching Summary**

### 6.13.2 ANALYSIS

A steady decline in average amount of time spent was noticeable over the last four phases, i.e. where a model was being used (except Phase 9). This decline was attributed to the steady refinement of the model, resulting in less and less time being required by the learners to find their information.

Note: Phase 9 consisted of three sessions, with a total learner count of only 22 – an average of only 7.3 learners per session. The overall average for the other 43 sessions (1087 learners) is 25.3 learners per session. The small sample size per session for Phase 9 could be the reason for the average time of this phase not following the expected trend in the graph above.

The author concluded that the model in its various forms had the overall effect of decreasing the time spent to find relevant information on the Internet.

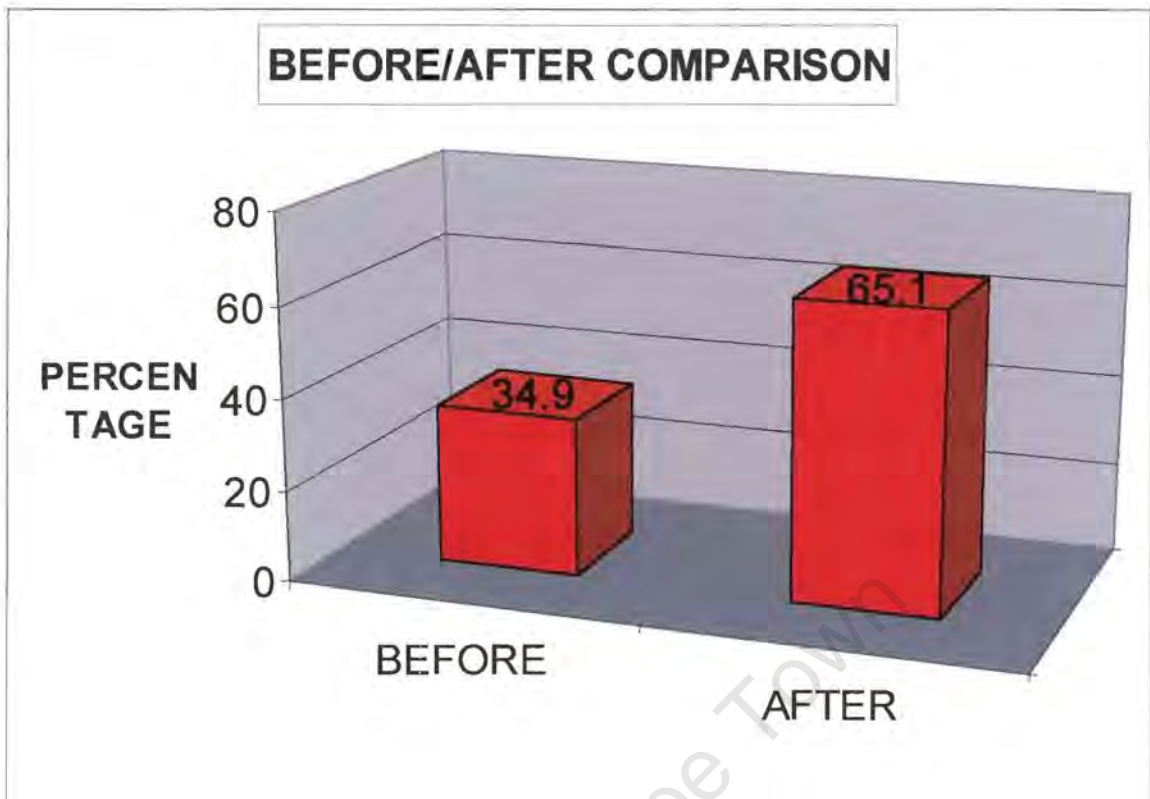
## 6.14 YES ANSWERS AFTER MODEL 1, 2 AND 3

### 6.14.1 RESULTS

The figures in Table 6.14 below were one of the most significant success indicators of the model, especially when compared to the number of Yes answers before the use of the model.

PHASE	NO	YES	UNKNOWN
7	10	45	0
9	5	15	0
11	84	106	6
13	31	76	16
<b>TOTAL</b>	130	242	22
<b>%</b>	<b>34.9</b>	<b>65.1</b>	<b>-</b>

**TABLE 6.14 Yes Answers after Model 1,2,3 Summary**



**GRAPH 6.12 – Comparison of Yes Answers**

#### 6.14.2 ANALYSIS

The success of the searching model can be expressed in two different ways.

##### Method 1:

The first one was to simply compare the number of Yes answers (in other words: "I found the information") before any model, to those after exposure to the models.

The percentage of Yes answers before the model for the last four sets of participants (see Table 6.14) was 34.9%. The average percentage of Yes answers after the exposure to the models (i.e. phase 7, 9, 11 and 13) was 65.1% (see Table 6.14). This meant that 65 out of every 100 learners claimed to have found one piece of academic information in 30 minutes, with the help of one of the versions of the model.

This represented an increase of 30.2%, which was considered to be substantial, and already suggested that the searching model was successful.

The second method will be explained in Section 6.16.2 below.

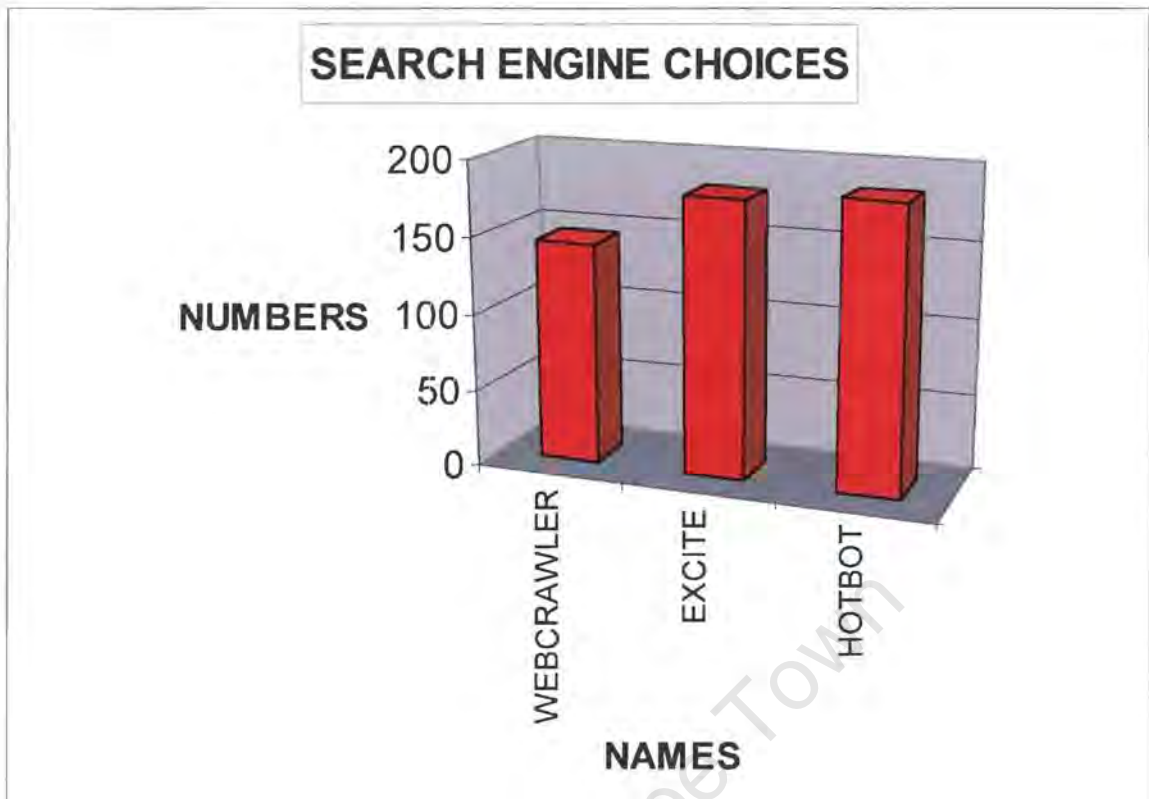
## **6.15        NUMBER OF USERS CHOOSING WEBCRAWLER/EXCITE/HOTBOT**

### **6.15.1      RESULTS**

The reason why this question was included, was to see if any bias in the choice of search engine, as recorded earlier with Yahoo!, (see Section 5.3.1) was still present.

<b>PHASE</b>	<b>WEBCRAWLER</b>	<b>EXCITE</b>	<b>HOTBOT</b>
<b>11</b>	67	97	110
<b>13</b>	78	83	75
<b>TOTAL</b>	145	180	185
<b>%</b>	<b>28.4</b>	<b>35.3</b>	<b>36.3</b>

**TABLE 6.15 – Number of Users choosing Search Engines Summary**



**GRAPH 6.13 – Number of Users choosing Search Engines Summary**

### 6.15.2 ANALYSIS

The ratio of choices between Webcrawler/Excite/Hotbot is 28.4/35.3/36.3. Compared to the wide differences reported earlier in the study (see Section 6.9.1), these differences are too small to be of consequence.

Statistical testing has proved that there is a strong association between the number of choices of each pair of search engines of 0.1%. Each one of the following was tested as a pair, thereby covering all three possible relations: Webcrawler and Excite, Webcrawler and Hotbot, and Excite and Hotbot (see Appendix M, Section 6.15.2).

The main reason for the even spread among the search engines was the limit of only three to choose from, compared to the unlimited choice before.

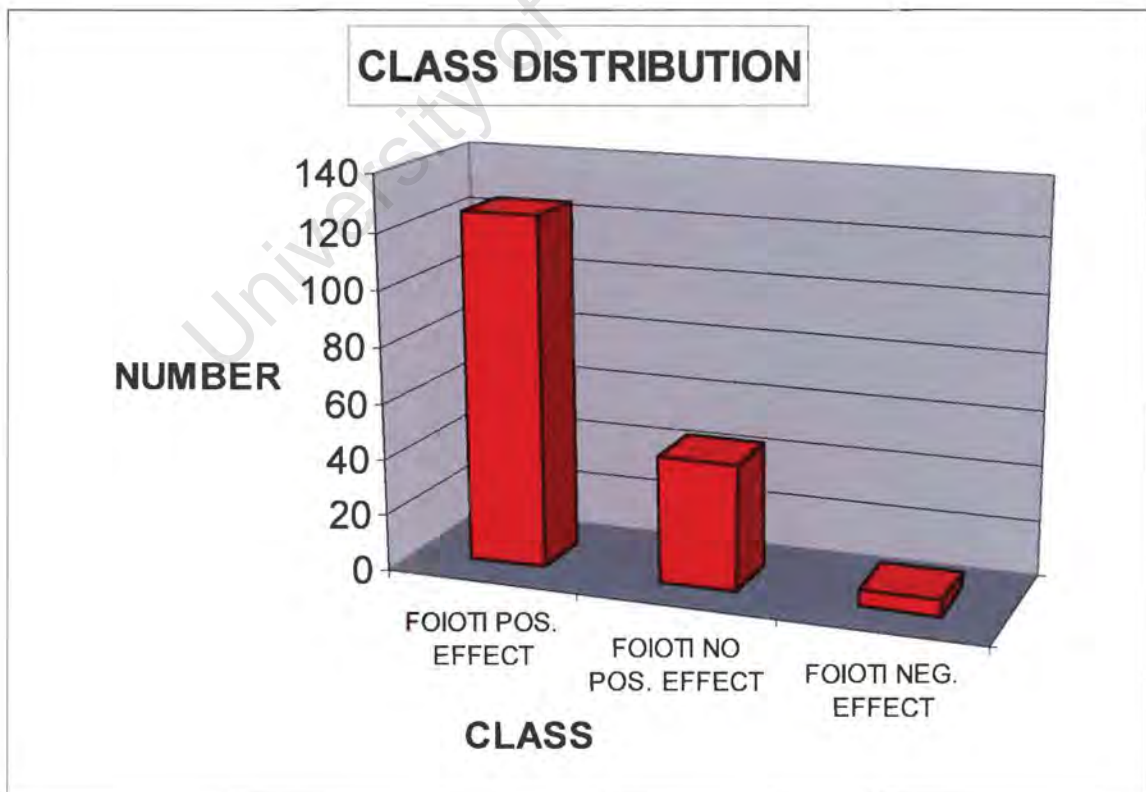
## 6.16 NUMBER OF PARTICIPANTS PER CLASS

### 6.16.1 RESULTS

The last two sets of experiments (phase 11 and 13) were the final tests to establish whether or not Model 3, FOIOTI, was successful. After the definition of the five classes, and the omission of the two non-performance-related ones (see Section 5.10), the following table summarizes the combined figures.

CLASS	NUMBER	PERCENTAGE
1	125	71.0
3	45	25.6
4	6	3.4
<b>TOTAL</b>	<b>176</b>	<b>100</b>

**TABLE 6.16 – Number of Participants per Class Summary**



**GRAPH 6.14 – Number of Users per Class – FOIOTI**

### 6.16.2 ANALYSIS

The second method of determining the model success (see Section 6.14.2 for a description of the first method) will now be discussed.

#### Method 2:

The second method was to consider only the experiments where a model was involved (phase 7, 9, 11 and 13), and compare the distribution of the different classes of possible answers. The reader is reminded of the definition of five different classes (see Section 5.10) of responses, varying from Class 1 (model had a positive effect) to Class 4 (model had a negative effect). Class 5 contained invalid responses. From Table 6.16 it was clear that 71.0% of the participants of the last two phases reported that FOIOTI had a positive effect on their searching ability (Class 1). Another 25.6% reported that FOIOTI had no positive effect (Class 3) and 3.4% noted that FOIOTI had a negative effect on their searching abilities (Class 4).

# **CHAPTER 7**

## **CONCLUSION**

### **7.1 INTRODUCTION**

The purpose of this chapter is to draw the major conclusions from this study. A summary of the salient points is given below.

### **7.2 SUMMARY**

#### **7.2.1 DEMOGRAPHICS**

A total of 1 109 IT/IS learners spread over three continents, 13 cities and 20 institutions, were personally supervised by the author in 46 separate sessions involving Internet searching during this study. The sessions varied between 45 minutes and 3 hours each. The average age of the participants of 22.2 years, their gender distribution of one female per 1.89 males and the representation of the four major race categories were supported by other authors as being statistically acceptable. A number of invalid responses were omitted, to guarantee the reliability of the data, while ensuring that the number of participants remaining was still well above the calculated minimum sample size. Finally, no single learner took part in this study twice. Obviously the same participant had to take part twice towards the end of the study, i.e. searching before the model and searching after the model at the same institute.

At this stage it was concluded that the demographics of this study meet all conditions required ensuring the accuracy of the results.

### 7.2.2 INTERNET EXPOSURE

The following was determined:

- Most learners had little or no official training on the use of the Internet (this was confirmed by Holland and Powell, who found that few learners at a university received formal instruction in the use of tools for electronic information retrieval (Holland and Powell 1995:7).
- A large percentage of these same learners used the Internet regularly.
- Approximately the same percentage also claimed that they had used search engines more than once before.
- Furthermore, a disproportionately large percentage of learners chose the same search engine when given a free choice.

These four facts seemed to indicate that most learners were using an IT tool regularly without having been trained, which is probably the reason why the success figures without assistance were so low.

It was therefore concluded that there was a definite need for the proposed searching model.

### 7.2.3 SEARCHING DETAIL

During the experiments it was confirmed that many users would choose the same, well-known search engine, probably as a result of following the example of a neighbour. The exceptionally large number of website summaries received in response to searching confirmed the suspicion that many learners were not using the features available to them to focus their searches. In all actual searches executed, with and without the model, the learners had to define their own query on an academic topic in their field. They were also instructed to use a different topic during the after-model search than during the pre-model search. The degree of difficulty of the search queries was therefore consistent.

However, the visible decrease in the number of results received over the last phases was an early indicator of the success of the models.

Even the number of websites visited and the time spent before achieving success decreased as the experiments progressed. The bias towards choosing certain search engines was removed by limiting the choice of search engines, at the same time giving guidance in an area previously neglected.

#### **7.2.4 SIGNIFICANCE**

At this stage the author would like to reflect on some issues mentioned at the start of the study, to be able to draw this project to a close.

The author is of the opinion that the regular use of FOIOTI could address and solve many of the problems identified in Chapter 1, as outlined below.

- Attention was drawn to the one-way communication and rote learning activities present at schooling level (see Section 1.2.1). Although the experiments in this study were done with higher education learners, the author has done tests with school children using FOIOTI (mostly grade 9, approximately 15 years of age). In all cases, positive results were found in terms of finding information required for school projects in History, Geography and Science. Learners other than those from tertiary institutions could use FOIOTI to assist them in the learning process.
- It was proved conclusively in this study that the casual use of standard search engines is not a reliable method to locate specific information on the Internet. Only 231 out of 717 learners, i.e. 32.2%, managed to locate one piece of specific information using standard search engines (see Section 6.12.2, second last paragraph). The OBE system is a resource-based system (see Section 1.2.4), and educators could use FOIOTI in a computer laboratory as a tool to tap the Internet for information.
- Problems were noted in motivating learners to put more into their studies (see Section 1.3), to participate in classes (See Section 1.4) and to assume responsibility for their own progress (see Section 1.6). When a learner is required to use FOIOTI to find his/her own reference materials, the first steps on the road to solving these problems are being taken. The two informal experiments (Section 1.4 – Participation, and Section 1.5.2 – Resource Location) serve as basis for this belief.

- Traditional resources, including textbooks, handouts and transparencies suffer from some serious drawbacks (see Sections 1.5.1.1 and 1.5.1.2). It was noted that a learning process took place on the educator's side while preparing learning materials such as transparencies (see Section 1.5.1.2), but that the learner did not inherit this automatically by taking ownership of the learning material. The author introduced his own learners to FOIOTI during Phase 13. It was clear that the learners were in fact enthused and inspired and, as a result, became part of the learning process while using FOIOTI.

A parallel was noticed between:

EDUCATOR	>	CREATE	>	LEARNING MATERIALS, and
LEARNER	>	FIND	>	RESOURCES.

In general, the problem of regular replacement of outdated textbooks was also addressed by the concept of using the Internet as information resource. If it could be assumed that the educator would evaluate websites before referring the learners to them, the validity of the information on websites of choice would be guaranteed. Therefore, the authors of the webpages automatically do the updating of the information base in question. This solved the problem of information in textbooks being out of date.

- The library appears to be an under-utilized resource (see Section 1.5.1.3). The Internet and libraries offer the same service in broad terms, namely that of being information sources. However, the information available in a library is organized in a way that makes retrieval of relevant information a relatively easy task. The Internet, in stark contrast, has no scheme of cataloguing data, which is one of the reasons why finding relevant information is so difficult (see Section 2.5.8). At this stage it is probably impossible for a library to attempt setting up a catalogue of Internet resources, owing to the sheer size, changing nature and unreliability of the Internet-based information database.

- Libraries at educational institutions could incorporate Internet services as part of their informational offering. A possible role of a library could be to assist its users to find answers to their queries by improving their search strategies, or simply finding the answers on behalf of the users. The author attempted this concept at his home institution, but a lack of manpower resources made the implementation difficult.
- The availability of Internet access in an educational institution computer laboratory could draw learner attention away from productive work. However, if resource hunting through the use of FOIOTI is incorporated as part of a subject, the effect of this phenomenon could be minimized.

Considering the changing environment in South Africa, and similar situations elsewhere, the results of this project could have far-reaching effects on the way an educator involves the learners in their own education process. Learners from disadvantaged backgrounds could now receive the benefit of saving textbook expenses, and should be able to work on their own to a larger extent than before, since the dependence on these textbooks can be decreased.

## **7.2.5 FURTHER RESEARCH**

Since no study can be fully exhaustive, the ideas below could be incorporated in other research projects.

### **7.2.5.1 FOCUSED FOIOTI**

The same concept of FOIOTI could be used to develop a similar product, aimed at a specific market, for example: the Engineering, Medical, Political or Academic fields. Databases, which specialize in one of these fields, can be identified, and used as the proposed search services.

### **7.2.5.2 HYPER HELP**

During the study the author was appalled by the low standard of the help menus of most of the search engines. Help menus were sometimes non-existent, and almost always difficult to find.

Navigation of most of these menus was non-intuitive, and even with a clear picture in mind of what a user was looking for (for example: listing of operators, meaning of terms, examples of queries) the user often had to give up in despair. This was confirmed by Winship (Winship 1995:51), and by the authors of a study of search agents (Tegenbos and Nieuwenhuysen 1997:140).

Especially noticeable was the absence of actual examples of how search operators could be used to focus a search. A few brief examples are listed below. In all five cases the link to the help menus on the home page was small and difficult to find.

#### AltaVista

Some reasonable examples were listed in the "Search Cheat Sheet". It is not the most likely place a novice user would look, however – there were no links to "Examples" or "Queries" or "Operator examples" from anywhere in the help pages.

#### Excite

Looking under "Search Help" and "Advanced Search Tips" yielded no examples. Only after looking under "General Search Tips", some basic examples were given.

#### Hotbot

Only some very basic examples on length of phrases and the omission of stop words were listed. No complex queries involving the combination of operators were supplied.

#### Infoseek

Only at the point when the user was prepared to drill down through a number of menu levels, did the help become accessible. Once again, just very basic examples were given.

#### SavvySearch

Once again only basic examples were shown, with no complex operators or combinations offered as example.

A program could be designed which is incorporated into the search box of a search engine.

When a user submits a query, this program can start a simulated conversation with the user, probing for essential components needed to construct a query with a reasonable chance of success. It can eventually propose a query which involves a number of operators, based on a built-in knowledge database. During this process, and from the proposed query with explanation, the user will learn how to construct sensible queries. An example of such a “conversation” between a fictional but typical search engine and a user follows. Read the table diagonally from left to right to left while moving from top to bottom.

University of Cape Town

SEARCH ENGINE DISPLAYS...	... USER TYPES IN AS ANSWER
What do you want to search for?	
	Pictures of cars
Do you need photographs or drawings of cars?	
	A photograph
What kind of car?	
	A BMW
What BMW model?	
	The Z3
<u>I suggest you type the following into the search box:</u>	
url:BMW AND "Z3" +.jpeg -.gif -"7-series" -"5-series" -"3-series"	

**Explanation:**

1. Since BMW is a well-known brandname, the company will probably have their own website. The operator "url" I have used above, searches for a website whose address includes the term "BMW".
2. You specified the model Z3. Since the capital letter has a specific meaning, I have put it in quotes, to ensure that z3 is not included. The AND word (in capitals) is a Boolean operator which ensures that only websites whose name includes BMW AND which refer to "Z3", are included in your answers.
3. The + operator, called inclusion, will ensure that the sites listed in your answers contain a reference to .jpeg, which is a type of file used when saving photographs.
4. The – operator will exclude any references to what follows the operator. Since you don't want photographs of other models, I have suppressed references to these model ranges. I have also suppressed references including .gif files, which are drawings as opposed to photographs.

**TABLE 7.1 – Hyper Help Example**

### 7.2.5.3 SPECIFICATION ENHANCER

The author has found in many years of exposure to learners, their way of learning and their problems, that few of them have the ability to express their thoughts, and specifically a search request, in a natural language. If a learner cannot express his/her thoughts in his/her home language (or at least in English), how can he/she expect to communicate these thoughts in a strange, computer-based language? This computer language can be programming in a standard language (COBOL, Visual Basic, C++, Java, etc) or search engine coding.

As an example, if a user cannot verbalize or type:

"I need to find the specifications of the new Campagnolo 10-speed racing bicycle rear cluster",

how can he/she be expected to progress to the construction of the query:

+Campag\* +("10-speed" OR "10 speed") +cluster +spec\* -price -shop -bargain?

The author considers the first step in searching (see the "Spec box" in FOIOTI) to be of paramount importance, namely writing down at least one full sentence stating exactly what the user is looking for. The way in which FOIOTI then extracts information from the user bears some resemblance to the work done by Austin on the PRECIS system (Austin 1984:9). In this case, the modes are reversed: an indexer has to evaluate a reference and generate a "subject statement", which is a "title-like" phrase. This statement was then analysed into separate terms, where the role of each term was considered. This role-searching involved finding answers to questions such as:

- What happened?
- To what (or whom) did it happen?
- In what location did it happen?

Even the final step in Austin's process, where a computer produced the required index entries, has a parallel in FOIOTI calling the search engine(s) to find the required information.

Without connecting each one of FOIOTI's boxes to a role, the user is actually forced to pull his/her own search specification apart until enough substantial meaning is found to fill each box as instructed by the screen messages.

A possible future project could be a natural language processing front end which requests a full sentence from the user, validates it to be more than a minimum length, ensures that it contains less than a certain percentage of stop-words (meaningless but commonly used words such as: and, the, that, a), etc. The program will then prod the user for more detail - from possibly a single word to a full sentence, which can then be submitted as the query.

### **7.2.6          SEARCHING SUCCESS**

A low percentage of learners found the information they had specified within the time limit and the original hypothesis was proved to be false, leading the researcher to consider the alternative hypothesis to be true. This was indeed proved by the results of searching experiments using FOIOTI.

### **7.3              CONCLUSION**

The author raised some issues regarding learner participation and resources, based on own experience and some experimentation with learners (see Section 1.7). He found that learners were hesitant to participate in a class situation, but eagerly took part in an Internet searching experiment, even though it involved academic matters. A question was raised about the feasibility of an educational tool or method which would increase learner involvement and motivation. Problems with the use of traditional resources were also noted.

It is believed that these issues are solved in a constructive way, by suggesting that FOIOTI be employed as a standard resource locating tool, regardless of the subject content.

Some problems were experienced during the searching experiments. A large amount of time and money was expended on travelling right through South Africa (on two occasions). Two overseas visits added to the expense.

In some cases, after establishing contact and making arrangements with colleagues at remote institutions, experiments were conducted in less than favourable circumstances. Outdated equipment, slow Internet connections (confirmed by: Hirsh 1999:1270, Voorbij 1999:604) and misunderstandings about the two separate sessions made it difficult to persevere with the experiments. In others, after spending two days at an institution, as few as 12 learners were seen. In one extreme case, the author had to leave an institution without seeing a single learner, after having confirmed all details more than once beforehand. In the process three days of travelling and some expenses were wasted.

In some minor ways the experiments failed to meet expectations. Some learners either refused or were unable to comply with the simple requirements of filling in the FOIOTI boxes as instructed. Sometimes the institution's timetable dictated that the experimental session be separated and offered over two days. This resulted in a break in concentration, and many invalid responses due to learners only attending one of the two sessions. Some learners could not type out a full specification for FOIOTI, resulting in "No" answers since other boxes were also incomplete. It was assumed that FOIOTI required a certain minimum level of communication skills in the English language. If those skills were lacking, FOIOTI did not deliver a positive result. However, currently the only way an Internet user can translate his/her perception of the information need into a query which a search engine can accept, is by way of the English language.

The experiment results could probably have been improved if a smaller sample had been covered. This way, more attention could have been given to the training sessions, possibly even covering language deficiencies.

However, the author is convinced that the following advice from a prominent IR researcher has been heeded, with positive results:

“The success or failure of any interactive system and technology is contingent on the extent to which user issues, the human factors, are addressed right from the beginning to the very end, right from theory, conceptualization, and design process on to development, evaluation, and to provision of services.”

(Saracevic 1999:1062).

This project has contributed to the existing body of knowledge in the fields of education and information technology. It provides a way for educators to guide their students in finding their own study materials, thereby involving them in their own progress. In a broader sense, the author found that FOIOTI is equally well suited for use by almost any average computer user. In this sense, “average computer user” refers to a user who is computer literate, and might be used to working on an Internet browser program. Previous exposure to search engines is not a prerequisite, although it would help merely because the user is then normally more familiar with the program interfaces, and is not deterred by the mass of information often produced by these programs. Finally, the results of this project provide a stepping stone for any learner who has to reposition himself/herself in terms of a new skill which can find use in any discipline: the efficient retrieval of relevant data from a rich data source.

### **FINAL CONCLUSIONS:**

**One of the two major success indicators is the increase in the number of learners who found their information – it increased from 32.2% (without help) to 65.1% by using FOIOTI.**

**The other major indicator of success is a comparison of the number of learners who managed to increase, experience no effect or decrease their searching abilities through the use of FOIOTI. A total of 71.0% reported that FOIOTI enabled them to find information whereas they were unsuccessful before, 25.6% reported no positive change and only 3.4% claimed that they could not find information with FOIOTI, but did find it without the model.**

Overall, the author has learned a tremendous amount during this project, and would like to recommend that colleagues consider using FOIOTI (<http://www.mwe.co.za>) to further their own learners' education in a practical and positive way.

University of Cape Town

# APPENDIX A

## Form 1

University of Cape Town

## INTERNET SEARCHING EXPERIMENT

<b>A</b>	<p>1 This exercise is NOT a test of your knowledge about the internet, or of your computer skills.</p> <p>2 Please complete block B below. All detail is confidential, and is only required to make it possible to summarize results.</p> <p>3 Finally work through block C, D and E, and supply the answers you <b>ACTUALLY FOUND</b>, not those you <b>THINK YOU SHOULD FIND</b>.</p>
----------	---

<b>B</b>	Initials: <input style="width: 90%;" type="text"/>	Surname: <input style="width: 90%;" type="text"/>	Date: <input style="width: 90%;" type="text"/>
	Age: <input style="width: 10%;" type="text"/> yrs	Institute: <input style="width: 80%;" type="text"/>	City: <input style="width: 80%;" type="text"/>
	Gender: <input type="checkbox"/> M <input type="checkbox"/> F	Race: <input type="checkbox"/> African <input type="checkbox"/> Asian <input type="checkbox"/> Coloured <input type="checkbox"/> White <input type="checkbox"/> Other:	
Gender and race are required for statistical purposes only.			
Course: <input style="width: 95%;" type="text"/>			
Subjects:	1	3	5
	2	4	6

<b>C</b>	Search Number: 1	What exactly are you going to search for:
how can I convert a given number from decimal to hexadecimal?		
Which search engine are you going to use: <input style="width: 80%;" type="text"/> http://www.		
What keyword or phrase are you going to specify: <input style="width: 95%;" type="text"/>		

<b>D</b>	How many results did the search engine supply: <input style="width: 80%;" type="text"/>
	How many of these sites did you actually visit: <input style="width: 80%;" type="text"/>
	How much time did you spend on this search: <input style="width: 80%;" type="text"/> minutes
	Did you find exactly what you were looking for? <input type="checkbox"/> Y <input type="checkbox"/> N

<b>E</b>	REMARKS
What was the biggest problem (if any) you have experienced while completing this form?	
Any other comments about this experiment?	

# APPENDIX B

## Form 2

University of Cape Town

## INTERNET SEARCHING EXPERIMENT

**A** This experiment attempts to determine how learners search for information on the internet. Wait for the lecturer to help you fill in the first page of this form. It is NOT a test of your knowledge or skills, and it is anonymous. Please supply the correct information!

**B** DATE     98 CITY  INSTITUTION

Course (write out in full)

Subjects	1 <input type="text"/>	3 <input type="text"/>	5 <input type="text"/>
(in full)	2 <input type="text"/>	4 <input type="text"/>	6 <input type="text"/>

The following information is for statistical purposes only, and will be treated as confidential.

Age  yrs Gender  M  F Race  African  Asian  Colored  White  Other

**C** How many hours of formal instruction have you received on how to use the internet?

1.  2.  3.

Please indicate how often you use the internet 4.

5.  6.  7.

**D** Identify any one subject (tick the numbered box - see B above) on which you would like to get more information

1	2	3	4	5	6
---	---	---	---	---	---

**E** Identify any one resource you are using in this subject, which does not supply all the information you need (fill in detail in any one block)

Handbook name	Description of notes or handouts	Library name	Video name	Computer manual name	Other

**F** Specify EXACTLY what you will be searching for which, if you find it, will supply the information you need (see block E above)


You have 30 minutes to work through this page at your own pace.

**G** Choose any one search engine which you would like to use, and fill its name in below.

http://www.

Now load this program

**H1** If the search engine you have chosen requires (a) keyword(s), specify what you are going to type in

Now type it in and proceed

**H2** If the search engine you have chosen requires successive choices of topics, make and list these choices now

1	3	5
2	4	6

**I** If you had to refine your search further, eg by searching inside results a second time, or by making further choices, please explain what you did

---



---



---

**J** How many results did the search engine supply by the time you have started investigating some websites?

How many of these websites did you actually visit?

**K** Read over your specification of what you were going to search for again - please see Block F on previous page.

Did you find EXACTLY what you were looking for?

 Y  N

Please note: there is no CORRECT or WRONG answer to this question - simply tick the box which is true in your case

How much time did you spend on this search?

 minutes

**L** Any comments?

---



---



---



---

# APPENDIX C

## Form 3

University of Cape Town

## INTERNET SEARCHING EXPERIMENT

**A** This experiment attempts to determine how learners search for information on the internet. It is NOT a test of your knowledge or skills, and it is ANONYMOUS. Wait for assistance before you fill in the first page of this form. Please supply the correct information!

**B** DATE 

a	d	m	m

 2 0 0 0 CITY  INSTITUTION

Course ( write out in full)

Subjects	1 <input style="width: 100%;" type="text"/>	3 <input style="width: 100%;" type="text"/>	5 <input style="width: 100%;" type="text"/>
(in full)	2 <input style="width: 100%;" type="text"/>	4 <input style="width: 100%;" type="text"/>	6 <input style="width: 100%;" type="text"/>

The following information is required for statistical purposes only.

Age  yrs Gender  M  F Race  African  Asian  Colored  White  Other

**C** How many hours of formal instruction have you received on the use of the internet?

1  0 to 1 hr      2  1+ to 5 hrs      3  5+ hrs

**D** How often do you use the internet?

1.  have never used it    2.  approx. once/mnth    3.  approx. once/week    4.  at least once/day

**E** How often have you used an internet search engine to search for information on the internet?

1  never      2  once only      3  more than once

**F** Identify any one subject (tick the numbered box - see B above) on which you would like to get more information

1    2    3    4    5    6

**G** Identify any one resource you are using in this subject (F above), which does not supply all the information you need (fill in detail in any one block). This MUST be based on block F.

Handbook name:	Description of notes or handouts:	Library name:	Video name:	Computer manual name:	Lecture topic:	Other

**H** Specify EXACTLY what you will be searching for which, if you find it, will supply the information you need (see block G above). This specification MUST be based on block G.


You have 30 minutes to work through this page at your own pace.

**I** Choose any one search engine which you would like to use, and fill its name in below.

http://www.

Now load this program

**J1** If the search engine you have chosen requires (a) keyword(s), specify EXACTLY what you are going to type in

Now type it in and proceed

**J2** If the search engine you have chosen requires successive choices of topics, make and list these choices now

1	3	5
2	4	6

**K** If you had to refine your search further, eg by searching inside results a second time, or by making further choices, please explain what you did


**L** How many results did the search engine supply by the time you have started investigating some websites?

How many of these websites did you actually visit?

**M** Read over your specification of what you were going to search for again - please see Block H on previous page.

Did you find EXACTLY what you were looking for?

 Y  N

Please note: there is no CORRECT or WRONG answer to this question - simply tick the box which is true in your case

How much time did you spend on this search?

 minutes

**N** Any comments?

# APPENDIX D

## Form 4

University of Cape Town

<b>INTERNET SEARCHING EXPERIMENT</b>		Code.						
<b>A</b> This experiment attempts to determine how learners search for information on the internet. It is NOT a test of your knowledge or skills, and it is ANONYMOUS. Wait for assistance before you fill in the first page of this form. Please supply the correct information!								
<b>B</b>	DATE <table border="1" style="display: inline-table; border-collapse: collapse;"><tr><td style="width: 20px; text-align: center;">d</td><td style="width: 20px; text-align: center;">d</td><td style="width: 20px; text-align: center;">m</td><td style="width: 20px; text-align: center;">m</td></tr></table> 2 0 0 0	d	d	m	m	CITY <input style="width: 100px;" type="text"/>		
d	d	m	m					
INSTITUTION <input style="width: 100%;" type="text"/>								
Course ( write out in full) <input style="width: 100%;" type="text"/>								
Subjects	<table border="1" style="display: inline-table; border-collapse: collapse;"><tr><td style="width: 30px; text-align: center;">1</td><td style="width: 30px; text-align: center;">3</td><td style="width: 30px; text-align: center;">5</td></tr></table>	1	3	5				
1	3	5						
(in full)	<table border="1" style="display: inline-table; border-collapse: collapse;"><tr><td style="width: 30px; text-align: center;">2</td><td style="width: 30px; text-align: center;">4</td><td style="width: 30px; text-align: center;">6</td></tr></table>	2	4	6				
2	4	6						
The following information is required for statistical purposes only.								
Age <input style="width: 30px;" type="text"/> yrs Gender <input type="checkbox"/> M <input type="checkbox"/> F Race <input type="checkbox"/> African <input type="checkbox"/> Asian <input type="checkbox"/> Colored <input type="checkbox"/> White <input type="checkbox"/> Other								
<b>C</b> How many hours of formal instruction have you received on the use of the internet?								
<table style="width: 100%;"><tr><td style="width: 33%; text-align: center;"><input type="checkbox"/> 1. 0 to 1 hr</td><td style="width: 33%; text-align: center;"><input type="checkbox"/> 2. 1+ to 5 hrs</td><td style="width: 33%; text-align: center;"><input type="checkbox"/> 3. 5+ hrs</td></tr></table>			<input type="checkbox"/> 1. 0 to 1 hr	<input type="checkbox"/> 2. 1+ to 5 hrs	<input type="checkbox"/> 3. 5+ hrs			
<input type="checkbox"/> 1. 0 to 1 hr	<input type="checkbox"/> 2. 1+ to 5 hrs	<input type="checkbox"/> 3. 5+ hrs						
<b>D</b> How often do you use the internet?								
<table style="width: 100%;"><tr><td style="width: 25%; text-align: center;"><input type="checkbox"/> 1. have never used it</td><td style="width: 25%; text-align: center;"><input type="checkbox"/> 2. approx. once/mnth</td><td style="width: 25%; text-align: center;"><input type="checkbox"/> 3. approx. once/week</td><td style="width: 25%; text-align: center;"><input type="checkbox"/> 4. at least once/day</td></tr></table>			<input type="checkbox"/> 1. have never used it	<input type="checkbox"/> 2. approx. once/mnth	<input type="checkbox"/> 3. approx. once/week	<input type="checkbox"/> 4. at least once/day		
<input type="checkbox"/> 1. have never used it	<input type="checkbox"/> 2. approx. once/mnth	<input type="checkbox"/> 3. approx. once/week	<input type="checkbox"/> 4. at least once/day					
<b>E</b> How often have you used an internet search engine to search for information on the internet?								
<table style="width: 100%;"><tr><td style="width: 33%; text-align: center;"><input type="checkbox"/> 1. never</td><td style="width: 33%; text-align: center;"><input type="checkbox"/> 2. once only</td><td style="width: 33%; text-align: center;"><input type="checkbox"/> 3. more than once</td></tr></table>			<input type="checkbox"/> 1. never	<input type="checkbox"/> 2. once only	<input type="checkbox"/> 3. more than once			
<input type="checkbox"/> 1. never	<input type="checkbox"/> 2. once only	<input type="checkbox"/> 3. more than once						
<b>F</b> Identify any one subject (tick the numbered box - see B above) on which you would like to get more information								
<table style="width: 100%;"><tr><td style="width: 16.6%; text-align: center;"><input type="checkbox"/> 1</td><td style="width: 16.6%; text-align: center;"><input type="checkbox"/> 2</td><td style="width: 16.6%; text-align: center;"><input type="checkbox"/> 3</td><td style="width: 16.6%; text-align: center;"><input type="checkbox"/> 4</td><td style="width: 16.6%; text-align: center;"><input type="checkbox"/> 5</td><td style="width: 16.6%; text-align: center;"><input type="checkbox"/> 6</td></tr></table>			<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6			
<b>G</b> Identify any one resource you are using in this subject (F above), which does not supply all the information you need (fill in detail in any one block). This MUST be based on block F.								
Handbook name:	Description of notes or handouts:	Library name:	Video name:	Computer manual name:	Lecture topic:	Other		
<b>H</b> Specify EXACTLY what you will be searching for which, if you find it, will supply the information you need (see block G above). This specification MUST be based on block G.								
<input style="width: 100%; height: 20px;" type="text"/>								
<input style="width: 100%; height: 20px;" type="text"/>								
<input style="width: 100%; height: 20px;" type="text"/>								

You have 30 minutes to work through this page at your own pace.		Code: <input style="width: 50px;" type="text"/>						
<b>I</b> Choose any one search engine which you would like to use, and fill its name in below.								
<input style="width: 95%; height: 20px;" type="text" value="http://www."/>								
Now load this program								
<b>J1</b> If the search engine you have chosen requires (a) keyword(s), specify EXACTLY what you are going to type in	<b>J2</b> If the search engine you have chosen requires successive choices of topics, make and list these choices now							
<input style="width: 95%; height: 30px;" type="text"/>	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 33%; text-align: center;">1</td> <td style="width: 33%; text-align: center;">3</td> <td style="width: 33%; text-align: center;">5</td> </tr> <tr> <td style="text-align: center;">2</td> <td style="text-align: center;">4</td> <td style="text-align: center;">6</td> </tr> </table>		1	3	5	2	4	6
1	3	5						
2	4	6						
Now type it in and proceed								
<b>K</b> If you had to refine your search further, eg by searching inside results a second time, or by making further choices, please explain what you did								
<b>L</b> How many results did the search engine supply by the time you have started investigating some websites? <input style="width: 50px;" type="text"/>								
How many of these websites did you actually visit? <input style="width: 50px;" type="text"/>								
<b>M</b> Read over your specification of what you were going to search for again - please see Block H on previous page.								
Did you find EXACTLY what you were looking for? <input style="width: 20px;" type="checkbox"/> Y <input style="width: 20px;" type="checkbox"/> N								
Please note: there is no CORRECT or WRONG answer to this question - simply tick the box which is true in your case								
How much time did you spend on this search? <input style="width: 50px;" type="text"/> minutes								
<b>N</b> Any comments?								

# APPENDIX E

## Form 5

University of Cape Town

<b>INTERNET SEARCH NUMBER 2</b>		Code:	
<b>A</b>	Did you use the FOIOTI program for this search?	<input type="checkbox"/> Y	<input type="checkbox"/> N
<a href="http://home.mweb.co.za/me/melius/seamodel.htm">http://home.mweb.co.za/me/melius/seamodel.htm</a>			
<b>B</b>	Please copy over your full search specification in this box		
<hr/> <hr/> <hr/> <hr/>			
<b>C</b>	Put a tick in the box of every FOIOTI search engine you have used during this searching session:		
<input type="checkbox"/> Webcrawler <input type="checkbox"/> Excite <input type="checkbox"/> Hotbot			
<b>D</b>	How many websites did you visit?	<input type="text"/>	
<b>E</b>	How much time did you spend on this search?	<input type="text"/>	min
<b>F</b>	Did you find what you were looking for (see Box B above)?	<input type="checkbox"/> Y	<input type="checkbox"/> N
<b>G</b>	Comments		
<hr/> <hr/> <hr/> <hr/>			

# APPENDIX F

## Model 1

University of Cape Town

<b>1</b>	<b>INTERNET SEARCHING</b>	This is a model you can use to help you find information on the internet. Please start at Box 2, and follow the arrows. Use ticks for Box 2 to 9 to show your choices.
Mr. Melius Weideman, Cape Town, South Africa meliusw@yahoo.com April 1999		
<b>2</b>	<b>PREREQUISITES</b>	a. Can you read English? <input type="checkbox"/> Y <input type="checkbox"/> N      b. Do you need to find information? <input type="checkbox"/> Y <input type="checkbox"/> N
		c. Do you have the necessary hardware, software and internet connection to search the internet? <input type="checkbox"/> Y <input type="checkbox"/> N
<b>3</b>	<b>TOPIC</b> Describe EXACTLY what you want to search for in at least one full English sentence (see Example 1).	

The flowchart starts with a decision diamond '5' asking 'Is your access to the internet slow?'. If 'NO', it goes to '6 CHOOSE PROGRAM'. If 'YES', it goes to '7 SLEEPING SERVERS'. From '7', it goes to '9 CHOOSE PROGRAM'. From '5', it also goes to '9'. From '9', it goes to '8 COUNTRY DEPENDANCE'. From '8', it goes to '9'. From '9', it goes to '10 BUILD KEYWORD QUERY'.

**6 CHOOSE PROGRAM**  
Tick any search program here which you want to use for your searching (see MAJOR PLAYERS)

AV	EX	GT
HB	IN	LO
LY	MG	MS
NL	NS	SN
YA	AI	AJ
DO	ME	MM
PR	SA	

**7 SLEEPING SERVERS**  
Tick any search program in box 9 which is currently sleeping, according to your time of day and location (see Example 2 and MAJOR PLAYERS)

**9 CHOOSE PROGRAM**

Australia:	AV-AU	EX-AU	YA-AU				
Asia:	AV-AS	United Kingdom:	EX-UK	LY-UK	UP		
					YA-UK		
Canada:	AV-CA	LY-CA	YA-CA				
USA:	AV	EX	GT	HB	IN	LO	LY
	MG	MS	NL	NS	SN	YA	AI
	AJ	DO	ME	MM	PR	SA	

**8 COUNTRY DEPENDANCE**  
Tick any search program which is in or close to the country of your choice in Box 9 below

<b>10</b>	<b>BUILD KEYWORD QUERY</b>	The hints below work for most of the MAJOR PLAYERS. Look at the HELP pages of the search engine you have chosen for more information:
<ul style="list-style-type: none"> <li>a. Use the + sign to specify words to be included (see Example 3)</li> <li>b. Use the - sign to specify words to be excluded (see Example 4)</li> <li>c. Use " " to specify order or capitalization (see Example 5)</li> <li>d. Limit range by domain (see Example 6)</li> <li>e. Limit range by URL (see Example 7)</li> <li>f. Limit range by what appears in web page title (see Example 8)</li> <li>g. Include other words by using wildcard * (see Example 9)</li> </ul>		
Read through your own TOPIC description in Box 3 again. Consider the KEYWORD Examples given in this Box (10). Now write out the KEYWORD QUERY below by using some of or combining a. to g. above - see Example 10. If you are unsure, use a few relevant, lowercase keywords without any symbols as your KEYWORD QUERY		

<b>11</b>	<b>EXECUTE SEARCH</b>	a. Load as many copies of the browser program of your choice as you have ticks in Box 6 or 9      b. Load the search program(s) you have ticked in box 6 or 9 above, one in each browser window
		c. Type in the keyword queries you have created in Box 10 above, and start the search (also time yourself from this point)
		d. Note down the first number of hits in Box 12 below, and attempt to narrow down the hits to between 1 and ±20
		e. Visit all the sites you consider necessary, and complete all the boxes below

<b>12</b> How many hits did the search engine supply the 1st time? <input type="text"/>	<b>13</b> How many hits did you have before visiting sites? <input type="text"/>	<b>14</b> How much time did you spend on this search? <input type="text"/> min.	<b>15</b> Did you find what you were looking for? <input type="checkbox"/> Y <input type="checkbox"/> N	Please write any comments on the back.
---	--	---	---	--

# APPENDIX G

## Model 2

University of Cape Town

## INTERNET SEARCHING

a model to help you  
find information on the Internet

- 1 **DEFINE** what you want to search for, by writing down at least one full English sentence (see Example 1). \_\_\_\_\_

- 2 **CHOOSE** between one and three of the Search Engines below, by putting a tick(s) in its box. See Major Players.

AV	HB	LY	NL	UP	AJ	MM
EX	IN	MG	NS	YA	DO	PR
GT	LO	MS	SN	AI	ME	SA

- 3 **FORMULATE** a Keyword Query based on your definition in Box 1 above. Use the hints below, and the online Help if necessary.

- Use the + sign to specify words to be included (see Example 2).
- Use the - sign to specify words to be excluded (see Example 3).
- Use " " to specify order and capitalization (see Example 4).
- Limit number of answers by domain (see Example 5).
- Limit number of answers by URL (see Example 6).
- Limit number of answers by what appears in the web page title (see Example 7).
- Include related words by using the wildcard \* (see Example 8).
- Combine any of the above to limit the number of answers (see Example 10).

Please spend a minute to: **read** through your definition in Box 1  
**apply** a. to h. to your definition.

Now **formulate** your Keyword Query and write it down (if unsure, write down a few related lowercase keywords from Box 1):

- 4 **EXECUTE** the search: load as many copies of your browser as you have ticks in Box 2. Now type in the URL's of each search engine you have chosen, one per browser. Type your Keyword Query (Box 3 above) into each Search Box, and start the search.

- 5 Please complete the boxes below.

How many sites did you visit?	How much time did you spend on this search? min.	Did you find what you were looking for?	Y	N
-------------------------------	--	---	---	---

Acer Africa Pty Ltd and



Partially sponsor this project. Melius Weideman, Cape Technikon, SOUTH AFRICA meliusw@yahoo.com

# APPENDIX H

## Screen dump of Model 3 (FOIOTI)

University of Cape Town



# FOIOTI

## Finder Of Information On The Internet.

This page was designed for use with Internet Explorer 4.0 and higher. Netscape users might find that the FOIOTI program consistently produces "no results" from the Search Engines, or that it is not visible as a grey block with white text boxes.

Simply read the instructions in the grey block below, and type in the requested information!

FOIOTI will produce a specification in Search Engine Language which has a good chance of helping you find the information you need. It will also load the Search Engine(s) you choose and start the search.

**PLEASE READ THE HELP PAGE FOR MORE INFORMATION!**

!!!HELP!!!!



!!!HELP!!!!

**Describe EXACTLY what you are looking for by typing one sentence in the spec box (a single word or even a short phrase is not good enough!)**

**Spec box**

**Type one word from the spec box which is normally speit with (a) capital(s)**

**Caps box**

**Type two or three words from the spec box, which have a special meaning if they are used together. (one space between each word)**

**Phrase box**

**Type two other words from the spec box (one word per box below), which describe your topic**

**Word boxes**

**ALL FIVE BOXES ABOVE MUST BE COMPLETED BEFORE PROCEEDING!!**

Select any Search Engine by clicking on one of the buttons below

Webcrawler

Excite

Hotbot

# APPENDIX I

Appendix I is not used to prevent confusion with the figure “1” in the text

University of Cape Town

# APPENDIX J

## Experiment Results

University of Cape Town

THIS APPENDIX CONTAINS THE ACTUAL RAW DATA FROM EACH PHASE OF ALL THE EXPERIMENTS.

**J.1      RESULTS AND ANALYSIS: PHASE 1 – DESIGN AND TEST**  
**INSTRUMENT: FORM 1**

**J.1.1      AGE**

	<b>AGE</b>
<b>Maximum</b>	45
<b>Average</b>	21.5
<b>Minimum</b>	18

**J.1.2      GENDER**

<b>GENDER</b>	<b>NUMBER</b>	<b>PERCENTAGE</b>
<b>Male</b>	74	70.5
<b>Female</b>	31	29.5
<b>Unknown</b>	0	0
<b>TOTAL</b>	<b>105</b>	<b>100</b>

**J.1.3      RACE**

<b>RACE</b>	<b>NUMBER</b>	<b>PERCENTAGE</b>
<b>White</b>	60	57.1
<b>Coloured</b>	35	33.3
<b>African</b>	5	4.8
<b>Asian</b>	5	4.8
<b>Other</b>	0	0
<b>Unknown</b>	0	0
<b>TOTAL</b>	<b>105</b>	<b>100</b>

## J.1.4 VALIDITY

VALIDITY	NUMBER	PERCENTAGE
Valid	102	97.1
Invalid	3	2.9
<b>TOTAL</b>	<b>105</b>	<b>100</b>

## J.1.5 SEARCH ENGINES

SEARCH ENGINE	NUMBER	PERCENTAGE
Yahoo!	66	64.7
Infoseek	13	12.8
Metacrawler	5	4.9
AltaVista	4	3.9
Unknown	3	2.9
Webcrawler	3	2.9
Hotbot	2	2.0
Netscape Search	2	2.0
Goto	1	1.0
Lycos	1	1.0
Not Done	1	1.0
Search	1	1.0
<b>TOTAL</b>	<b>102</b>	<b>100</b>

## J.1.6 RESULTS RECEIVED

	RESULTS RECEIVED
Maximum	14 955 214
Average	584 233
Minimum	0

**J.1.7 NUMBER OF WEBSITES VISITED**

	WEBSITES VISITED
Maximum	19
Average	2.5
Minimum	0

**J.1.8 NUMBER OF YES ANSWERS BEFORE MODEL**

ANSWER	NUMBER	PERCENTAGE
No	63	61.8
Yes	34	33.3
Unknown	5	4.9
TOTAL:	102	100

**J.1.9 TIME SPENT SEARCHING**

	TIME SPENT
Maximum	30
Average	11.1
Minimum	0

**J.2 RESULTS AND ANALYSIS: PHASE 3 – TEST FORM 2 AS INSTRUMENT**

**J.2.1 AGE**

	AGE
Maximum	25
Average	20.6
Minimum	19

## J.2.2 GENDER

GENDER	NUMBER	PERCENTAGE
Male	18	94.7
Female	1	5.3
Unknown	0	0
<b>TOTAL</b>	<b>19</b>	<b>100</b>

## J.2.3 RACE

RACE	NUMBER	PERCENTAGE
White	18	94.7
Coloured	1	5.3
African	0	0
Asian	0	0
Other	0	0
Unknown	0	0
<b>TOTAL</b>	<b>19</b>	<b>100</b>

## J.2.4 VALIDITY

VALIDITY	NUMBER	PERCENTAGE
Valid	17	89.5
Invalid	2	10.5
<b>TOTAL</b>	<b>19</b>	<b>100</b>

## J.2.5 FORMAL INSTRUCTION

F. INSTR.	NUMBER	PERCENTAGE
1	16	94.1
3	1	5.9
2	0	0
Unknown	0	0
<b>TOTAL</b>	<b>17</b>	<b>100</b>

**J.2.6 INTERNET USAGE**

I. USAGE	NUMBER	PERCENTAGE
4	10	58.8
3	4	23.5
1	2	11.8
2	1	5.9
Unknown	0	0
<b>TOTAL</b>	<b>17</b>	<b>100</b>

**J.2.7 SEARCH ENGINES**

SEARCH ENGINE	NUMBER	PERCENTAGE
Yahoo!	7	41.2
Hotbot	4	23.5
AltaVista	2	11.7
Excite	1	5.9
Infoseek	1	5.9
Unknown	1	5.9
Webcrawler	1	5.9
<b>TOTAL</b>	<b>17</b>	<b>100</b>

**J.2.8 RESULTS RECEIVED**

	RESULTS RECEIVED
<b>Maximum</b>	132 542
<b>Average</b>	7 852
<b>Minimum</b>	1

**J.2.9 NUMBER OF WEBSITES VISITED**

	WEBSITES VISITED
Maximum	7
Average	4.3
Minimum	0

**J.2.10 NUMBER OF YES ANSWERS BEFORE MODEL**

ANSWER	NUMBER	PERCENTAGE
No	8	47
Yes	8	47
Unknown	1	6
TOTAL	17	100

**J.2.11 TIME SPENT SEARCHING**

	TIME SPENT
Maximum	30
Average	19.7
Minimum	5

**J.3 RESULTS AND ANALYSIS: PHASE 5 – TEST FORM 3**

**J.3.1 AGE**

	AGE
Maximum	42
Average	21.6
Minimum	17

## J.3.2 GENDER

GENDER	NUMBER	PERCENTAGE
Male	191	65.2
Female	91	31.1
Unknown	11	3.7
<b>TOTAL</b>	<b>293</b>	<b>100</b>

## J.3.3 RACE

RACE	NUMBER	PERCENTAGE
African	138	47.1
White	110	37.6
Asian	30	10.2
Coloured	8	2.7
Unknown	7	2.4
Other	0	0
<b>TOTAL</b>	<b>293</b>	<b>100</b>

## J.3.4 VALIDITY

VALIDITY	NUMBER	PERCENTAGE
Valid	236	80.5
Invalid	57	19.5
<b>TOTAL</b>	<b>293</b>	<b>100</b>

## J.3.5 FORMAL INSTRUCTION

F. INSTR.	NUMBER	PERCENTAGE
1	177	75
2	39	16.5
3	20	8.5
Unknown	0	0
<b>TOTAL</b>	<b>236</b>	<b>100</b>

**J.3.6 INTERNET USAGE**

<b>I. USAGE</b>	<b>NUMBER</b>	<b>PERCENTAGE</b>
4	120	50.9
3	81	34.3
2	25	10.6
1	10	4.2
Unknown	0	0
<b>TOTAL</b>	<b>236</b>	<b>100</b>

**J.3.7 SEARCH ENGINE USAGE**

<b>SE. USAGE</b>	<b>NUMBER</b>	<b>PERCENTAGE</b>
3	193	81.8
1	30	12.7
2	13	5.5
Unknown	0	0
<b>TOTAL</b>	<b>236</b>	<b>100</b>

## J.3.8 SEARCH ENGINES

SEARCH ENGINE	NUMBER	PERCENTAGE
Yahoo!	120	50.9
AltaVista	23	9.8
Webcrawler	22	9.3
Infoseek	18	7.6
Ananzi	10	4.2
Not Done	9	3.9
Excite	8	3.4
Search	5	2.1
Hotbot	4	1.7
Lycos	3	1.3
Zazoo	3	1.3
Microsoft	2	0.9
Aardvark	1	0.4
Goto	1	0.4
IBM	1	0.4
Mamma	1	0.4
Metacrawler	1	0.4
Netscape Search	1	0.4
Unknown	1	0.4
Webferret	1	0.4
Webopedia	1	0.4
<b>TOTAL</b>	<b>236</b>	<b>100</b>

## J.3.9 RESULTS RECEIVED

	RESULTS RECEIVED
Maximum	53 469 078
Average	606 348
Minimum	0

## J.3.10 NUMBER OF WEBSITES VISITED

	WEBSITES VISITED
Maximum	42
Average	4.0
Minimum	0

## J.3.11 NUMBER OF YES ANSWERS BEFORE MODEL

ANSWER	NUMBER	PERCENTAGE
No	135	57.2
Yes	90	38.1
Unknown	11	4.7
TOTAL	236	100

## J.3.12 TIME SPENT SEARCHING

	TIME SPENT
Maximum	30
Average	20.7
Minimum	1

J.4 RESULTS AND ANALYSIS: PHASE 7 – TEST FORM 3 WITH MODEL 1

## J.4.1 AGE

	AGE
Maximum	43
Average	24.1
Minimum	19

## J.4.2 GENDER

GENDER	NUMBER	PERCENTAGE
Male	50	73.5
Female	15	22.1
Unknown	3	4.4
<b>TOTAL</b>	<b>68</b>	<b>100</b>

## J.4.3 RACE

RACE	NUMBER	PERCENTAGE
White	63	92.7
Coloured	2	2.9
Unknown	2	2.9
African	1	1.5
Asian	0	0
Other	0	0
<b>TOTAL</b>	<b>68</b>	<b>100</b>

## J.4.4 VALIDITY

VALIDITY	NUMBER	PERCENTAGE
Valid	55	80.9
Invalid	13	19.1
<b>TOTAL</b>	<b>68</b>	<b>100</b>

## J.4.5 FORMAL INSTRUCTION

F. INSTR.	NUMBER	PERCENTAGE
1	43	78.2
2	11	20.0
3	1	1.8
Unknown	0	0.0
<b>TOTAL</b>	<b>55</b>	<b>100</b>

**J.4.6 INTERNET USAGE**

I. USAGE	NUMBER	PERCENTAGE
4	25	45.4
3	15	27.3
2	11	20.0
1	4	7.3
Unknown	0	0
<b>TOTAL</b>	<b>55</b>	<b>100</b>

**J.4.7 SEARCH ENGINE USAGE**

SE. USAGE	NUMBER	PERCENTAGE
3	49	89.1
1	4	7.3
2	2	3.6
Unknown	0	0.0
<b>TOTAL</b>	<b>55</b>	<b>100</b>

**J.4.8 SEARCH ENGINES**

SEARCH ENGINE	NUMBER	PERCENTAGE
Yahoo!	28	50.9
AltaVista	17	30.9
Hotbot	4	7.4
Lycos	2	3.6
Google	2	3.6
Metacrawler	1	1.8
Microsoft	1	1.8
<b>TOTAL</b>	<b>55</b>	<b>100</b>

## J.4.9 RESULTS RECEIVED

	RESULTS RECEIVED
Maximum	1 589 213
Average	72 295
Minimum	1

## J.4.10 NUMBER OF WEBSITES VISITED

	WEBSITES VISITED
Maximum	43
Average	6.7
Minimum	1

## J.4.11 NUMBER OF YES ANSWERS BEFORE MODEL

ANSWER	NUMBER	PERCENTAGE
No	36	65.5
Yes	19	34.5
Unknown	0	0.0
TOTAL	55	100

## J.4.12 TIME SPENT SEARCHING

	TIME SPENT
Maximum	30
Average	21.4
Minimum	5

**J.4.13 NUMBER OF YES ANSWERS AFTER MODEL 1**

ANSWER	NUMBER	PERCENTAGE
Yes	45	81.8
No	10	18.2
Unknown	0	0.0
<b>TOTAL</b>	<b>55</b>	<b>100</b>

**J.5 RESULTS AND ANALYSIS: PHASE 9 – TEST FORM 3 WITH MODEL 2****J.5.1 AGE**

	AGE
Maximum	39
Average	27.8
Minimum	21

**J.5.2 GENDER**

GENDER	NUMBER	PERCENTAGE
Female	12	54.5
Male	10	45.5
Unknown	0	0.0
<b>TOTAL</b>	<b>22</b>	<b>100</b>

**J.5.3 RACE**

RACE	NUMBER	PERCENTAGE
White	11	50.0
Asian	6	27.3
Other	3	13.7
Unknown	1	4.5
African	1	4.5
Coloured	0	0.0
<b>TOTAL</b>	<b>22</b>	<b>100</b>

**J.5.4 VALIDITY**

VALIDITY	NUMBER	PERCENTAGE
Valid	20	90.9
Invalid	2	9.1
<b>TOTAL</b>	<b>22</b>	<b>100</b>

**J.5.5 FORMAL INSTRUCTION**

F. INSTR.	NUMBER	PERCENTAGE
1	16	80.0
2	2	10.0
3	1	5.0
Unknown	1	5.0
<b>TOTAL</b>	<b>20</b>	<b>100</b>

**J.5.6 INTERNET USAGE**

I. USAGE	NUMBER	PERCENTAGE
4	17	85.0
3	3	15.0
2	0	0.0
1	0	0.0
Unknown	0	0.0
<b>TOTAL</b>	<b>20</b>	<b>100</b>

**J.5.7 SEARCH ENGINE USAGE**

I. USAGE	NUMBER	PERCENTAGE
3	19	95.0
Unknown	1	5.0
1	0	0.0
2	0	0.0
<b>TOTAL</b>	<b>20</b>	<b>100</b>

**J.5.8 SEARCH ENGINES**

SEARCH ENGINE	NUMBER	PERCENTAGE
AltaVista	5	25
Yahoo!	5	25
Webcrawler	4	20
Infoseek	2	10
AskJeeves	1	5
Lycos	1	5
Netscape	1	5
Snap	1	5
<b>TOTAL</b>	<b>20</b>	<b>100</b>

**J.5.9 RESULTS RECEIVED**

	RESULTS RECEIVED
Maximum	59 510 332
Average	3 241 376
Minimum	1

**J.5.10 NUMBER OF WEBSITES VISITED**

	WEBSITES VISITED
Maximum	30
Average	5.8
Minimum	0

**J.5.11 NUMBER OF YES ANSWERS BEFORE MODEL**

ANSWER	NUMBER	PERCENTAGE
No	12	60.0
Yes	8	40.0
Unknown	0	0.0
TOTAL	20	100

**J.5.12 TIME SPENT SEARCHING**

	TIME SPENT
Maximum	30
Average	19
Minimum	5

**J.5.13 NUMBER OF YES ANSWERS AFTER MODEL 2**

ANSWER	NUMBER	PERCENTAGE
Yes	15	75.0
No	5	25.0
Unknown	0	0.0
<b>TOTAL</b>	<b>20</b>	<b>100</b>

**J.6 RESULTS AND ANALYSIS: PHASE 11 – TEST FORM 4 WITH MODEL 3****J.6.1 AGE**

	AGE
<b>Maximum</b>	58
<b>Average</b>	22.6
<b>Minimum</b>	17

**J.6.2 GENDER**

GENDER	NUMBER	PERCENTAGE
Male	162	50.6
Female	131	40.9
Unknown	27	8.5
<b>TOTAL</b>	<b>320</b>	<b>100</b>

## J.6.3 RACE

RACE	NUMBER	PERCENTAGE
African	215	67.2
White	43	13.5
Unknown	32	10.0
Asian	19	5.9
Coloured	10	3.1
Other	1	0.3
<b>TOTAL</b>	<b>320</b>	<b>100</b>

## J.6.4 VALIDITY

VALIDITY	NUMBER	PERCENTAGE
Valid	196	61.3
Invalid	124	38.7
<b>TOTAL</b>	<b>320</b>	<b>100</b>

## J.6.5 FORMAL INSTRUCTION

F. INSTR.	NUMBER	PERCENTAGE
1	142	72.5
2	30	15.3
3	12	6.1
Unknown	12	6.1
<b>TOTAL</b>	<b>196</b>	<b>100</b>

**J.6.6 INTERNET USAGE**

<b>I. USAGE</b>	<b>NUMBER</b>	<b>PERCENTAGE</b>
<b>4</b>	<b>85</b>	<b>43.4</b>
<b>3</b>	<b>44</b>	<b>22.5</b>
<b>1</b>	<b>33</b>	<b>16.8</b>
<b>2</b>	<b>22</b>	<b>11.2</b>
<b>Unknown</b>	<b>12</b>	<b>6.1</b>
<b>TOTAL</b>	<b>196</b>	<b>100</b>

**J.6.7 SEARCH ENGINE USAGE**

<b>SE. USAGE</b>	<b>NUMBER</b>	<b>PERCENTAGE</b>
<b>3</b>	<b>102</b>	<b>52.0</b>
<b>1</b>	<b>69</b>	<b>35.2</b>
<b>2</b>	<b>15</b>	<b>7.7</b>
<b>Unknown</b>	<b>10</b>	<b>5.1</b>
<b>TOTAL</b>	<b>196</b>	<b>100</b>

**J.6.8 SEARCH ENGINES**

SEARCH ENGINE	NUMBER	PERCENTAGE
Unknown	88	44.9
Yahoo!	55	28.1
AltaVista	12	6.2
Infoseek	9	4.6
Webcrawler	9	4.6
Excite	6	3.1
37	3	1.5
Hotbot	3	1.5
Lycos	3	1.5
Looksmart	2	1.0
AllTheWeb	1	0.5
Ananzi	1	0.5
Cyber4	1	0.5
Goto	1	0.5
Highway 61	1	0.5
Netscape	1	0.5
<b>TOTAL</b>	<b>196</b>	<b>100</b>

**J.6.9 RESULTS RECEIVED**

	RESULTS RECEIVED
<b>Maximum</b>	48 337 250
<b>Average</b>	751 831.8
<b>Minimum</b>	0

**J.6.10 NUMBER OF WEBSITES VISITED**

	WEBSITES VISITED
Maximum	25
Average	3.3
Minimum	0

**J.6.11 NUMBER OF YES ANSWERS BEFORE MODEL 3**

ANSWER	NUMBER	PERCENTAGE
No	153	78.1
Yes	40	20.4
Unknown	3	1.5
<b>TOTAL</b>	<b>196</b>	<b>100</b>

**J.6.12 TIME SPENT SEARCHING**

	TIME SPENT
Maximum	30
Average	18.5
Minimum	0

**J.6.13 NUMBER OF USERS CHOOSING WEBCRAWLER**

ANSWER	NUMBER	PERCENTAGE
No	71	36.2
Yes	67	34.2
Unknown	58	29.6
<b>TOTAL</b>	<b>196</b>	<b>100</b>

**J.6.14 NUMBER OF USERS CHOOSING EXCITE**

ANSWER	NUMBER	PERCENTAGE
Yes	97	49.5
Unknown	58	29.6
No	41	20.9
<b>TOTAL</b>	<b>196</b>	<b>100</b>

**J.6.15 NUMBER OF USERS CHOOSING HOTBOT**

ANSWER	NUMBER	PERCENTAGE
Yes	110	56.1
Unknown	58	29.6
No	28	14.3
<b>TOTAL</b>	<b>196</b>	<b>100</b>

**J.6.16 NUMBER OF WEBSITES VISITED**

	NUMBER
Maximum	23
Average	3.2
Minimum	0

**J.6.17 AMOUNT OF TIME SPENT**

	TIME
Maximum	30
Average	19.8
Minimum	0

**J.6.18 NUMBER OF YES ANSWERS AFTER MODEL 3**

ANSWER	NUMBER	PERCENTAGE
Yes	106	54.1
No	84	42.8
Unknown	6	3.1
<b>TOTAL</b>	<b>196</b>	<b>100</b>

**J.6.19 NUMBER OF PARTICIPANTS PER CLASS**

ANSWER	NUMBER	PERCENTAGE
1	76	38.8
5	63	32.2
3	31	15.8
2	23	11.7
4	3	1.5
<b>TOTAL</b>	<b>196</b>	<b>100</b>

**J.6.20 NUMBER OF PARTICIPANTS PER CLASS (ADJUSTED - 5)**

ANSWER	NUMBER	PERCENTAGE
1	76	57.1
3	31	23.3
2	23	17.3
4	3	2.3
<b>TOTAL</b>	<b>133</b>	<b>100</b>

**J.6.21 NUMBER OF PARTICIPANTS PER CLASS (ADJUSTED - 2)**

ANSWER	NUMBER	PERCENTAGE
1	76	69.1
3	31	28.2
4	3	2.7
<b>TOTAL</b>	<b>110</b>	<b>100</b>

**J.7 RESULTS AND ANALYSIS: PHASE 13 – IMPLEMENT MODEL 3.****J.7.1 AGE**

	<b>AGE</b>
<b>Maximum</b>	27
<b>Average</b>	19.0
<b>Minimum</b>	17

**J.7.2 GENDER**

<b>GENDER</b>	<b>NUMBER</b>	<b>PERCENTAGE</b>
Male	131	46.5
Unknown	95	33.7
Female	56	19.8
<b>TOTAL</b>	<b>282</b>	<b>100</b>

**J.7.3 RACE**

<b>RACE</b>	<b>NUMBER</b>	<b>PERCENTAGE</b>
White	98	34.8
Unknown	98	34.8
Coloured	66	23.4
African	13	4.6
Asian	5	1.7
Other	2	0.7
<b>TOTAL</b>	<b>282</b>	<b>100</b>

**J.7.4 VALIDITY**

<b>VALIDITY</b>	<b>NUMBER</b>	<b>PERCENTAGE</b>
Invalid	159	56.4
Valid	123	43.6
<b>TOTAL</b>	<b>282</b>	<b>100</b>

**J.7.5 FORMAL INSTRUCTION**

<b>F. INSTR.</b>	<b>NUMBER</b>	<b>PERCENTAGE</b>
<b>1</b>	<b>97</b>	<b>78.9</b>
<b>Unknown</b>	<b>13</b>	<b>10.6</b>
<b>2</b>	<b>11</b>	<b>8.9</b>
<b>3</b>	<b>2</b>	<b>1.6</b>
<b>TOTAL</b>	<b>123</b>	<b>100</b>

**J.7.6 INTERNET USAGE**

<b>I. USAGE</b>	<b>NUMBER</b>	<b>PERCENTAGE</b>
<b>3</b>	<b>42</b>	<b>34.1</b>
<b>2</b>	<b>30</b>	<b>24.4</b>
<b>4</b>	<b>20</b>	<b>16.3</b>
<b>1</b>	<b>19</b>	<b>15.4</b>
<b>Unknown</b>	<b>12</b>	<b>9.8</b>
<b>TOTAL</b>	<b>123</b>	<b>100</b>

**J.7.7 SEARCH ENGINE USAGE**

<b>SE. USAGE</b>	<b>NUMBER</b>	<b>PERCENTAGE</b>
<b>3</b>	<b>74</b>	<b>60.2</b>
<b>1</b>	<b>28</b>	<b>22.8</b>
<b>Unknown</b>	<b>12</b>	<b>9.7</b>
<b>2</b>	<b>9</b>	<b>7.3</b>
<b>TOTAL</b>	<b>123</b>	<b>100</b>

**J.7.8 SEARCH ENGINES**

SEARCH ENGINE	NUMBER	PERCENTAGE
Unknown	53	43.0
Yahoo!	40	32.5
AltaVista	8	6.5
Infoseek	4	3.3
Lycos	4	3.3
Hotbot	3	2.5
Google	3	2.5
Webcrawler	2	1.6
Excite	2	1.6
Search	2	1.6
Ananzi	1	0.8
Netscape	1	0.8
<b>TOTAL</b>	<b>123</b>	<b>100</b>

**J.7.9 RESULTS RECEIVED**

	RESULTS RECEIVED
Maximum	2 107 225
Average	58 539.7
Minimum	1

**J.7.10 NUMBER OF WEBSITES VISITED**

	WEBSITES VISITED
Maximum	15
Average	3.3
Minimum	0

**J.7.11 NUMBER OF YES ANSWERS BEFORE MODEL 3**

ANSWER	NUMBER	PERCENTAGE
No	79	64.2
Yes	32	26.0
Unknown	12	9.8
<b>TOTAL</b>	<b>123</b>	<b>100</b>

**J.7.12 TIME SPENT SEARCHING**

	TIME SPENT
Maximum	30
Average	13.5
Minimum	0

**J.7.13 NUMBER OF USERS CHOOSING WEBCRAWLER**

ANSWER	NUMBER	PERCENTAGE
Yes	78	63.4
No	29	23.6
Unknown	16	13.0
<b>TOTAL</b>	<b>123</b>	<b>100</b>

**J.7.14 NUMBER OF USERS CHOOSING EXCITE**

ANSWER	NUMBER	PERCENTAGE
Yes	83	67.5
No	24	19.5
Unknown	16	13.0
<b>TOTAL</b>	<b>123</b>	<b>100</b>

**J.7.15 NUMBER OF USERS CHOOSING HOTBOT**

ANSWER	NUMBER	PERCENTAGE
Yes	75	61.0
No	32	26.0
Unknown	16	13.0
<b>TOTAL</b>	<b>123</b>	<b>100</b>

**J.7.16 NUMBER OF WEBSITES VISITED**

	NUMBER
Maximum	20
Average	3.6
Minimum	0

**J.7.17 AMOUNT OF TIME SPENT**

	TIME
Maximum	30
Average	19.4
Minimum	1

**J.7.18 NUMBER OF YES ANSWERS AFTER MODEL 3**

ANSWER	NUMBER	PERCENTAGE
Yes	76	61.8
No	31	25.2
Unknown	16	13.0
<b>TOTAL</b>	<b>123</b>	<b>100</b>

**J.7.19 NUMBER OF PARTICIPANTS PER CLASS**

ANSWER	NUMBER	PERCENTAGE
1	49	39.9
5	39	31.7
2	18	14.6
3	14	11.4
4	3	2.4
<b>TOTAL</b>	<b>123</b>	<b>100</b>

**J.7.20 NUMBER OF PARTICIPANTS PER CLASS (ADJUSTED - 5)**

ANSWER	NUMBER	PERCENTAGE
1	49	58.3
2	18	21.4
3	14	16.7
4	3	3.6
<b>TOTAL</b>	<b>84</b>	<b>100</b>

**J.7.21 NUMBER OF PARTICIPANTS PER CLASS (ADJUSTED - 2)**

ANSWER	NUMBER	PERCENTAGE
1	49	74.3
3	14	21.2
4	3	4.5
<b>TOTAL</b>	<b>66</b>	<b>100</b>

# APPENDIX K

## Examples

University of Cape Town

## Example 1

**Microsoft  
Office**



I need to find out how to set up a columnar report from a table using Microsoft Access 97, which is part of Microsoft Office.



**Unix**



I need to download a copy of a technical manual explaining how SCO Unix Version 7.0 uses semaphores to achieve mutual exclusion.



**C++ arrays**



I need the C++ source code of the fastest algorithm that will find the k'th biggest element in an array .



## Example 2-5

### Example 2: + Inclusion

visual basic tutorial      This KEYWORD SPECIFICATION will return hits about a programming language, but also about eyesight and all kinds of tutorials.

+visual +basic +tutorial      This should only return hits about VB tutorials.

### Example 3: - Exclusion

windows      This KEYWORD SPECIFICATION should return hits about Microsoft's operating system, but also about those things we use to look through walls.

+windows +Microsoft -3.1 -95 -98      This should only return hits about Windows NT.

### Example 4: " " Order/Capitalization

NeXT      This KEYWORD SPECIFICATION will return a large number of hits, only some of them about a type of computer.

"NeXT computer"      This should force both the order and capitalization, and cut out the non-computer based hits.

### Example 5: Domain/Host/Site (au, de, es, fr, it, mu, nl, uk, za...)

domain:si      This KEYWORD SPECIFICATION should return a large number of sites hosted in Slovenia.

domain:org      This should return a large number of sites which are organizations.

## *Example 6-9*

### **Example 6: URL**

url:ac.za	This KEYWORD SPECIFICATION should return a large number of academic sites hosted in South Africa.
url:intel	This should return sites which are strongly linked to Intel, including their own official site.

### **Example 7: Title**

title:"Systems Analysis"	This KEYWORD SPECIFICATION should return webpages whose title include these 2 words - this implies that the hits listed are most likely really about Systems Analysis.
--------------------------	--

### **Example 8: \* Expansion**

"Cobol cod*"	This KEYWORD SPECIFICATION should return some websites including references to Cobol code, Cobol coding, Cobol code sheets and Cobol coding errors.
--------------	---

### **Example 9: Combinations**

+computers +"programming code" -Cobol -Basic -Fortran -Java -Assembler -Pascal  
 This KEYWORD SPECIFICATION should return some hits containing examples of C++ coded programs, while filtering out the other languages mentioned.

"PC Operating Systems" -windows -microsoft  
 This KEYWORD SPECIFICATION should return sites about PC operating systems other than Microsoft Windows.

Intel +domains:uk  
 This KEYWORD SPECIFICATION should return sites containing the word Intel, which are hosted in the UK.

Acer Africa Pty Ltd and



Partially sponsor this project. Melius Weideman,  
 Cape Technikon, SOUTH AFRICA meliusw@yahoo.com

# APPENDIX L

## Source Code of Model 3 (FOIOTI)

University of Cape Town

```
// Decompiled by Decafe PRO - Java Decompiler
// Classes: 1  Methods: 38  Fields: 24
```

```
package Applets;
```

```
import java.applet.Applet;
import java.applet.AppletContext;
import java.awt.*;
import java.awt.event.ActionEvent;
import java.awt.event.ActionListener;
import java.beans.Beans;
import java.io.IOException;
import java.io.PrintStream;
import java.net.MalformedURLException;
import java.net.URL;
import java.util.EventObject;
```

```
public class SearchBox extends Applet
```

```
    implements ActionListener
```

```
{
```

```
private Button ivjButton1;
private Button ivjButton2;
private Button ivjButton3;
private Textfield ivjTextfield1;
private Textfield ivjTextfield2;
private Textfield ivjTextfield3;
private Label ivjLabel1;
private Label ivjLabel2;
private Label ivjLabel3;
private Label ivjLabel4;
private Label ivjLabel5;
private Label ivjLabel6;
private Label ivjLabel7;
private Label ivjLabel8;
private Label ivjLabel10;
private Label ivjLabel11;
private Label ivjLabel9;
private Textfield ivjTextfield4;
private Textfield ivjTextfield5;
private Label ivjLabel12;
private Label ivjLabel13;
private Label ivjLabel14;
private Label ivjLabel15;
private Label ivjLabel16;
```

```
public void actionPerformed (ActionEvent e)
```

```
{
```

```
    if (e.getSource() == getButton1())
        connEtoC1(e);
```



```

{
    URL url = new URL("http://www.webcrawler.com/cgi-
bin/WebWuery?search=%2B%22" + s1 + "%22%2B%22" + s2 + "%22%2B" +
s3 + "%2B" + s4 + "&src=wc_results&showSummary=false");
    url.openConnection();
    getAppletContext().showDocument(url, "_blank");
}
catch(MalformedURLException _ex) {}
catch(IOException +ex) {}
}

private void connEtoC1(ActionEvent arg1)
{
    try
    {
        button1_ActionPerformed(arg1);
    }
    catch (Throwable ivjExc)
    {
        handleException(ivjExc);
    }
}

private void connEtoC2(ActionEvent arg1)
{
    try
    {
        button2_ActionPerformed(arg1);
    }
    catch (Throwable ivjExc)
    {
        handleException(ivjExc);
    }
}

private void connEtoC3(ActionEvent arg1)
{
    try
    {
        button3_ActionPerformed(arg1);
    }
    catch (Throwable ivjExc)
    {
        handleException(ivjExc);
    }
}

public String editString(String s)
{

```

```

StringBuffer sb = new StringBuffer();
for(int l = 0; l < s.length(); l++)
    if(s.charAt(l) == '+')
    {
        sb.append('%');
        sb.append('2');
        sb.append('B');
    }
    else
    {
        sb.append(s.charAt(l));
    }

return sb.toString();
}

public String getAppletInfo()
{
    return "Applets.SearchBox created using VisualAge for Java.";
}

private Button getButton1()
{
    if(ivjButton1 == null)
        try
        {
            ivjButton1 = new Button();
            ivjButton1.setName("Button1");
            ivjButton1.setFont(new Font("dialog", 1, 14));
            ivjButton1.setBounds(549, 388, 112, 40);
            ivjButton1.setLabel("Hotbot");
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjButton1;
}

private Button getButton2()
{
    if(ivjButton2 == null)
        try
        {
            ivjButton2 = new Button();
            ivjButton2.setName("Button2");
            ivjButton2.setFont(new Font("dialog", 1, 14));
            ivjButton2.setBounds(303, 388, 112, 40);
            ivjButton2.setLabel("Excite");
        }
}

```

```

        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    }
    return ivjButton2;
}

private Button getButton3()
{
    if(ivjButton3 == null)
        try
        {
            ivjButton3 = new Button();
            ivjButton3.setName("Button3");
            ivjButton3.setFont(new Font("dialog", 1, 14));
            ivjButton3.setBounds(53, 388, 112, 40);
            ivjButton3.setLabel("Webcrawler");
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    }
    return ivjButton3;
}

private Label getLabel1()
{
    if(ivjLabel1 == null)
        try
        {
            ivjLabel1 = new Label();
            ivjLabel1.setName("Label1");
            ivjLabel1.setFont(new Font("dialog", 1, 14));
            ivjLabel1.setText("Describe EXACTLY what you are
looking for by typing one sentence in the spec box");
            ivjLabel1.setBounds(14, 10, 627, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    }
    return ivjLabel1;
}

private Label getLabel10()
{
    if(ivjLabel10 == null)
        try
        {

```

```

        ivjLabel10 = new Label();
        ivjLabel10.setName("Label10");
        ivjLabel10.setFont(new Font("dialog", 1, 14));
        ivjLabel10.setText("←----- will provide fewer results");
        ivjLabel10.setBounds(57, 447, 270, 23);
    }
    catch (Throwable ivjExc)
    {
        handleException(ivjExc);
    }
    return ivjLabel10;
}

private Label getLabel11()
{
    if(ivjLabel11 == null)
        try
        {
            ivjLabel11 = new Label();
            ivjLabel11.setName("Label11");
            ivjLabel11.setFont(new Font("dialog", 1, 14));
            ivjLabel11.setText("will provide more results ----- >");
            ivjLabel11.setBounds(424, 447, 243, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjLabel11;
}

private Label getLabel12()
{
    if(ivjLabel12 == null)
        try
        {
            ivjLabel12 = new Label();
            ivjLabel12.setName("Label12");
            ivjLabel12.setFont(new Font("dialog", 1, 14));
            ivjLabel12.setText("Spec box");
            ivjLabel12.setBounds(611, 56, 96, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjLabel12;
}

private Label getLabel13()

```

```

{
    if(ivjLabel13 == null)
        try
        {
            ivjLabel13 = new Label();
            ivjLabel13.setName("Label13");
            ivjLabel13.setFont(new Font("dialog", 1, 14));
            ivjLabel13.setText("Caps box");
            ivjLabel13.setBounds(611, 96, 96, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjLabel13;
}

private Label getLabel14()
{
    if(ivjLabel14 == null)
        try
        {
            ivjLabel14 = new Label();
            ivjLabel14.setName("Label14");
            ivjLabel14.setFont(new Font("dialog", 1, 14));
            ivjLabel14.setText("Phrase box");
            ivjLabel14.setBounds(611, 162, 96, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjLabel14;
}

private Label getLabel15()
{
    if(ivjLabel15 == null)
        try
        {
            ivjLabel15 = new Label();
            ivjLabel15.setName("Label15");
            ivjLabel15.setFont(new Font("dialog", 3, 14));
            ivjLabel15.setText("ALL FIVE BOXES ABOVE MUST BE
COMPLETED BEFORE PROCEEDING!!");
            ivjLabel15.setBounds(122, 319, 540, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
}

```

```

    }
    return ivjLabel15;
}

private Label getLabel16()
{
    if(ivjLabel16 == null)
        try
        {
            ivjLabel16 = new Label();
            ivjLabel16.setName("Label16");
            ivjLabel16.setFont(new Font("dialog", 1, 14));
            ivjLabel16.setText("Word boxes");
            ivjLabel16.setBounds(611, 273, 96, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjLabel16;
}

private Label getLabel2()
{
    if(ivjLabel2 == null)
        try
        {
            ivjLabel2 = new Label();
            ivjLabel2.setName("Label2");
            ivjLabel2.setFont(new Font("dialog", 1, 14));
            ivjLabel2.setText("(a single word or even a short phrase is not
            goor enough!);
            ivjLabel2.setBounds(14, 28, 617, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjLabel2;
}

private Label getLabel3()
{
    if(ivjLabel3 == null)
        try
        {
            ivjLabel3 = new Label();
            ivjLabel3.setName("Label3");
            ivjLabel3.setFont(new Font("dialog", 1, 14));

```

```

        ivjLabel3.setText("Type one word from the spec box which is
    );
    ivjLabel3.setBounds(14, 97, 336, 23);
    }
    catch (Throwable ivjExc)
    {
        handleException(ivjExc);
    }
    return ivjLabel3;
}

private Label getLabel4()
{
    if(ivjLabel4 == null)
        try
        {
            ivjLabel4 = new Label();
            ivjLabel4.setName("Label4");
            ivjLabel4.setFont(new Font("dialog", 1, 14));
            ivjLabel4.setText("normally spelt with () capital(s)");
            ivjLabel4.setBounds(14, 115, 334, 26);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjLabel4;
}

private Label getLabel5()
{
    if(ivjLabel5 == null)
        try
        {
            ivjLabel5 = new Label();
            ivjLabel5.setName("Label5");
            ivjLabel5.setFont(new Font("dialog", 1, 14));
            ivjLabel5.setText("Type two or three words from the spec
    box,");
            ivjLabel5.setBounds(14, 158, 342, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjLabel5;
}

private Label getLabel6()
{

```

```

if(ivjLabel6 == null)
    try
    {
        ivjLabel6 = new Label();
        ivjLabel6.setName("Label6");
        ivjLabel6.setFont(new Font("dialog", 1, 14));
        ivjLabel6.setText("which have a special meaning if they");
        ivjLabel6.setBounds(14, 177, 336, 23);
    }
    catch (Throwable ivjExc)
    {
        handleException(ivjExc);
    }
return ivjLabel6;
}

private Label getLabel7()
{
    if(ivjLabel7 == null)
        try
        {
            ivjLabel7 = new Label();
            ivjLabel7.setName("Label7");
            ivjLabel7.setFont(new Font("dialog", 1, 14));
            ivjLabel7.setText("are used together. (one space between
each word)");
            ivjLabel7.setBounds(14, 194, 382, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
return ivjLabel7;
}

private Label getLabel8()
{
    if(ivjLabel8 == null)
        try
        {
            ivjLabel8 = new Label();
            ivjLabel8.setName("Label8");
            ivjLabel8.setFont(new Font("dialog", 1, 14));
            ivjLabel8.setText("Type two other words from the spec box
(one word per box below), which describe your topic");
            ivjLabel8.setBounds(11, 237, 676, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
}

```

```

    }
    return ivjLabel8;
}

private Label getLabel9()
{
    if(ivjLabel9 == null)
        try
        {
            ivjLabel9 = new Label();
            ivjLabel9.setName("Label9");
            ivjLabel9.setFont(new Font("dialog", 1, 14));
            ivjLabel9.setText("Select any Search Engine by clicking on one
of the buttons below");
            ivjLabel9.setBounds(135, 352, 491, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjLabel9;
}

private TextField getTextField1()
{
    if(ivjTextField1 == null)
        try
        {
            ivjTextField1 = new TextField();
            ivjTextField1.setName("TextField1");
            ivjTextField1.setBounds(12, 56, 592, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjTextField1;
}

private TextField getTextField2()
{
    if(ivjTextField2 == null)
        try
        {
            ivjTextField2 = new TextField();
            ivjTextField2.setName("TextField2");
            ivjTextField2.setBounds(345, 97, 259, 23);
        }
        catch (Throwable ivjExc)
        {

```

```
        handleException(ivjExc);
    }
    return ivjTextField2;
}

private TextField getTextField3()
{
    if(ivjTextField3 == null)
        try
        {
            ivjTextField3 = new TextField();
            ivjTextField3.setName("TextField3");
            ivjTextField3.setBounds(345, 161, 259, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjTextField3;
}

private TextField getTextField4()
{
    if(ivjTextField4 == null)
        try
        {
            ivjTextField4 = new TextField();
            ivjTextField4.setName("TextField4");
            ivjTextField4.setBounds(12, 275, 131, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
    return ivjTextField4;
}

private TextField getTextField5()
{
    if(ivjTextField5 == null)
        try
        {
            ivjTextField5 = new TextField();
            ivjTextField5.setName("TextField5");
            ivjTextField5.setBounds(182, 275, 131, 23);
        }
        catch (Throwable ivjExc)
        {
            handleException(ivjExc);
        }
}
```

```

        return ivjTextField5;
    }

    private void handleException(Throwable throwable)
    {
    }

    public void init()
    {
        super.init();
        try
        {
            setName("SearchBox");
            setLayout(null);
            setSize(715, 490);
            add(getButton1(), getButton1().getName());
            add(getButton2(), getButton2().getName());
            add(getButton3(), getButton3().getName());
            add(getTextField1(), getTextField1().getName());
            add(getTextField2(), getTextField2().getName());
            add(getTextField3(), getTextField3().getName());
            add(getLabel1(), getLabel1().getName());
            add(getLabel2(), getLabel2().getName());
            add(getLabel3(), getLabel3().getName());
            add(getLabel4(), getLabel4().getName());
            add(getLabel5(), getLabel5().getName());
            add(getLabel6(), getLabel6().getName());
            add(getLabel7(), getLabel7().getName());
            add(getLabel8(), getLabel8().getName());
            add(getTextField4(), getTextField4().getName());
            add(getTextField5(), getTextField5().getName());
            add(getLabel9(), getLabel9().getName());
            add(getLabel10(), getLabel10().getName());
            add(getLabel11(), getLabel11().getName());
            add(getLabel12(), getLabel12().getName());
            add(getLabel13(), getLabel13().getName());
            add(getLabel14(), getLabel14().getName());
            add(getLabel15(), getLabel15().getName());
            add(getLabel16(), getLabel16().getName());
            initConnections();
        }
        catch(Throwable ivjExc)
        {
        }
    }

    private void initConnections()
    {
        getButton1().addActionListener(this);
        getButton2().addActionListener(this);
        getButton3().addActionListener(this);
    }

```

```

}

public static void main(String args[])
{
    try
    {
        Frame frame
        try
        {
            Class aFrameClass =
Class.forName("com.ibm.uvm.abt.edit.TestFrame");
            frame = (Frame)aFrameClass.newInstance();
        }
catch(Throwable _ex)
{

frame = new Frame();
}
Class iiCls = Class.forName("Applets.SearchBox");
ClassLoader iiClsLoader = iiCls.getClassLoader();
SearchBox aSearchBox = (SearchBox)Beans.instantiate(iiClsLoader,
"Applets.SearchBox")
frame.add("Center", aSearchBox);
frame.setSize(aSearchBox.getSize());
frame.setVisible(true);
}
catch(Throwable exception)
{
    System.err.println("Exception occurred in main() of java.applet.Applet");
exception.printStackTrace(System.out);
}
}

public SearchBox()
{
    ivjButton1 = null;
    ivjButton2 = null;
    ivjButton3 = null;
    ivjTextField1 = null;
    ivjTextField2 = null;
    ivjTextField3 = null;
    ivjLabel1 = null;
    ivjLabel2 = null;
    ivjLabel3 = null;
    ivjLabel4 = null;
    ivjLabel5 = null;
    ivjLabel6 = null;
    ivjLabel7 = null;
    ivjLabel8 = null;
    ivjLabel10 = null;
}

```

```
ivjLabel11 = null;  
ivjLabel9 = null;  
ivjTextField4 = null;  
ivjTextField5 = null;  
ivjLabel12 = null;  
ivjLabel13 = null;  
ivjLabel14 = null;  
ivjLabel15 = null;  
ivjLabel16 = null;  
}  
}
```

# APPENDIX M

## Statistical References

University of Cape Town

## Section 6.2.2

PHASE	1	3	5	7	9	11	13
n	100	17	228	55	19	176	107
STD. DEV.	3.7	1.4	2.3	6.0	4.0	4.9	1.7
AVERAGE	21.5	20.6	21.1	24.5	27.2	22.3	19.3
CV	17.2	6.8	11.1	24.5	14.9	22.1	8.7

### Ages of Participants (years)

## Section 6.7.2

NOTE: Copyright (c) 1989-1996 by SAS Institute Inc., Cary, NC, USA.  
 NOTE: SAS (r) Proprietary Software Release 6.12 TS060  
 Licensed to CAPE TECHNIKON, Site 0082144001.

### Formal Internet Instruction vs Internet Usage 1

#### Formal Internet Instruction

FORINS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0 - 1 Hours	491	79.1	491	79.1
1 +- 5 Hours	93	15.0	584	94.0
5 + Hours	37	6.0	621	100.0

Frequency Missing = 128

Chi-Square Test for Equal Proportions

-----  
 Statistic = 592.039    DF = 2    Prob = **0.001**

WARNING: 17% of the data are missing.

**Formal Internet Instruction vs Internet Usage****2**

## Internet Usage

INTUSE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Never	68	10.9	68	10.9
1/Month	89	14.3	157	25.2
1/Week	189	30.3	346	55.5
1/Day	277	44.5	623	100.0

Frequency Missing = 126

## Chi-Square Test for Equal Proportions

Statistic = 179.536 DF = 3 Prob = 0.001

WARNING: 17% of the data are missing.

**Formal Internet Instruction vs Internet Usage****3**

## TABLE OF FORINS BY INTUSE

FORINS(Formal Internet Instruction) INTUSE(Internet Usage)

Frequency Percent Row Pct Col Pct	Never	1/Month	1/Week	1/Day	Total
0-1 Hours	63 10.21 12.91 96.92	78 12.64 15.98 87.64	142 23.01 29.10 75.94	205 33.23 42.01 74.28	488 79.09
1 + -5 hours	2 0.32 2.17 3.08	11 1.78 11.96 12.36	36 5.83 39.13 19.25	43 6.97 46.74 15.58	92 14.91
5 + Hours	0 0.00 0.00 0.00	0 0.00 0.00 0.00	9 1.46 24.32 4.81	28 4.54 75.68 10.14	37 6.00
Total	65 10.53	89 14.42	187 30.31	276 44.73	617 100.00

Frequency Missing = 132

## STATISTICS FOR TABLE OF FORINS BY INTUSE

Statistic	DF	Value	Prob
Chi-Square	6	31.011	<b>0.001</b>

Effective Sample Size = 617

Frequency Missing = 132

WARNING: 18% of the data are missing.

## Section 6.9.2

NOTE: Copyright (c) 1989-1996 by SAS Institute Inc., Cary, NC, USA.

NOTE: SAS (r) Proprietary Software Release 6.12 TS060

Licensed to CAPE TECHNIKON, Site 0082144001.

Yahoo -- Search Engine

4

SRCEN	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Did not choose Yahoo!	271	45.7	271	45.7
Chose Yahoo!	322	54.3	593	100.0

Frequency Missing = 156

### Chi-Square Test for Equal Proportions

Statistic = 4.386    DF = 1    Prob = 0.036

i.e. significant at the 0.05 level of significance.

WARNING: 21% of the data are missing.

## Section 6.10.2

NOTE: Copyright (c) 1989-1996 by SAS Institute Inc., Cary, NC, USA.

NOTE: SAS (r) Proprietary Software Release 6.12 TS060

Licensed to CAPE TECHNIKON, Site 0082144001.

Number of Results - RESU

14

### NPAR1WAY PROCEDURE

Analysis of Variance for Variable RESU  
Classified by Variable PHASE\_NO

PHASE_NO	N	Mean	Among MS	Within MS
1	96	590318.38	3.040249E13	1.806679E13
3	17	7852.41		
5	198	600223.61	F-value	Prob > F
7	50	72295.22	1.683	<b>0.1229</b>
9	20	3241376.35		<b>Not Significant</b>
11	99	751831.77		
13	71	58539.72		

Average Scores Were Used for Ties

Number of Results - Phase 9 vs Phase 11

37

### NPAR1WAY PROCEDURE

Analysis of Variance for Variable RESU  
Classified by Variable PHASE\_NO

PHASE_NO	N	Mean	Among MS	Within MS
9	20	3241376.35	1.031236E14	5.015175E13
11	99	751831.77		
			F-value	Prob > F
			2.056	<b>0.1543</b>
				<b>Not Significant</b>

Average Scores Were Used for Ties

Number of Results - Phase 11 vs Phase 13

25

### NPAR1WAY PROCEDURE

Analysis of Variance for Variable RESU  
Classified by Variable PHASE\_NO

PHASE_NO	N	Mean	Among MS	Within MS
11	99	751831.768	1.987362E13	1.49851E13
13	71	58539.718		
			F-value	Prob > F
			1.326	<b>0.2511</b>
				<b>Not Significant</b>

Average Scores Were Used for Ties

Number of Results - Phase 9 vs Phase 13

49

Analysis of Variance for Variable RESU  
Classified by Variable PHASE\_NO

PHASE_NO	N	Mean	Among MS	Within MS
9	20	3241376.35	1.580795E14	3.776579E13
13	71	58539.72		
			F-value	Prob > F
			4.186	<b>0.0437</b>
				<b>Significant at the 0.05 level</b>

Average Scores Were Used for Ties

## Section 6.15.2

NOTE: Copyright (c) 1989-1996 by SAS Institute Inc., Cary, NC, USA.

NOTE: SAS (r) Proprietary Software Release 6.12 TS060

Licensed to CAPE TECHNIKON, Site 0082144001.

## WebCrawler vs Excite vs HotBot

5

FOIWC	FOIEX	FOIHB	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	0	0	195	44.4	195	44.4
0	0	1	45	10.3	240	54.7
0	1	0	32	7.3	272	62.0
0	1	1	22	5.0	294	67.0
1	0	0	16	3.6	310	70.6
1	0	1	3	0.7	313	71.3
1	1	0	11	2.5	324	73.8
1	1	1	115	26.2	439	100.0

Frequency Missing = 310

## WebCrawler vs Excite 19:13 Friday, August 4, 2000 8

## TABLE OF FOIWC BY FOIEX

FOIWC(foiWebCrawler)

FOIEX(foiExcite)

Frequency Percent Row Pct Col Pct	0	1	Total
0	240 54.67 81.63 92.66	54 12.30 18.37 30.00	294 66.97
1	19 4.33 13.10 7.34	126 28.70 86.90 70.00	145 33.03
Total	259 59.00	180 41.00	439 100.00

Frequency Missing = 310

## STATISTICS FOR TABLE OF FOIWC BY FOIEX

Statistic	DF	Value	Prob
Chi-Square	1	188.521	0.001

**Very significant, at the level of 0,001. (0,1 %), i.e. There is a significant difference in the choice between Webcrawler and Excite.**

**Webcrawler vs HotBot 19:13 Friday, August 4, 2000**

**TABLE OF FOIWC BY FOIHB**  
FOIWC(foiWebCrawler) FOIHB(foiHotBot)

Frequency Percent Row Pct Col Pct	0	1	Total
0	227 51.71 77.21 89.37	67 15.26 22.79 36.22	294 66.97
1	27 6.15 18.62 10.63	118 26.88 81.38 63.78	145 33.03
Total	254 57.86	185 42.14	439 100.00

Frequency Missing = 310

**STATISTICS FOR TABLE OF FOIWC BY FOIHB**

Statistic	DF	Value	Prob
Chi-Square	1	136.718	<b>0.001</b>

**Very significant, at the level of 0,001. (0,1 %)**

**i.e. There is a significant difference in the choice between Webcrawler and HotBot.**

Excite vs HotBot 19:13 Friday, August 4, 2000 12

TABLE OF FOIEX BY FOIHB  
FOIEX(foiExcite) FOIHB(foiHotBot)

Frequency Percent Row Pct Col Pct	0	1	Total
0	211 48.06 81.47 83.07	48 10.93 18.53 25.95	259 59.00
1	43 9.79 23.89 16.93	137 31.21 76.11 74.05	180 41.00
Total	254 57.86	185 42.14	439 100.00

Frequency Missing = 310

Frequency Missing = 310

WARNING: 41% of the data are missing.

STATISTICS FOR TABLE OF FOIEX BY FOIHB

Statistic	DF	Value	Prob
Chi-Square	1	144.394	0.001

Very significant, at the level of 0,001. (0,1 %)

i.e. There is a significant difference in the choice between Excite and HotBot.

# BIBLIOGRAPHY

University of Cape Town

**SEARCH ENGINE HOME PAGES: STANDARD SERVERS**

37.COM. 2001. Home page. [Online]. Available: <http://www.37.com>. [Site visited on 27/03/2001].

ALL-IN-ONE. 2000. Home page. [Online]. Available: <http://www.albany.net/allinone/> [Site visited on 14/03/2000].

ALTAVISTA. 2000. Home page. [Online]. Available: <http://www.altavista.com>. [Site visited on 01/07/2000].

ASKJEEVES. 2000. Home page. [Online]. Available: <http://www.askjeeves.com>. [Site visited on 23/06/2000].

DOGPILE. 2000. Home page. [Online]. Available: <http://www.dogpile.com>. [Site visited on 11/06/2000].

EXCITE. 2000. Home page. [Online]. Available: <http://www.excite.com>. [Site visited on 27/06/2000].

GOTO. 2000. Home page. [Online]. Available: <http://www.goto.com>. [Site visited on 25/06/2000].

HOTBOT. 2001. Home page. [Online]. Available: <http://www.hotbot.com>. [Site visited on 27/03/2001].

INFOSEEK. 2000. Home page. [Online]. Available: <http://www.infoseek.com>. [Site visited on 17/06/2000].

LOOKSMART. 2000. Home page. [Online]. Available: <http://www.looksmart.com>. [Site visited on 05/06/2000].

LYCOS. 2000. Home page. [Online]. Available: <http://www.lycos.com>. [Site visited on 28/05/2000].

MAGELLAN. 2000. Home page. [Online]. Available: <http://www.mckinley.com>. [Site visited on 12/04/2000].

MAMMA. 2000. Home page. [Online]. Available: <http://www.mamma.com>. [Site visited on 09/06/2000].

METACRAWLER. 2000. Home page. [Online]. Available: <http://www.metacrawler.com>. [Site visited on 14/06/2000].

METASEARCH. 2000. Home page. [Online]. Available: <http://www.metasearch.com>. [Site visited on 15/05/2000].

NETSCAPE SEARCH. 2000. Home page. [Online]. Available: <http://www.netscape.com>. [Site visited on 17/05/2000].

NORTHERN LIGHT. 2000. Home page. [Online]. Available: <http://www.nlsearch.com>. [Site visited on 29/06/2000].

OPENTEXT. 2001. Home page. [Online]. Available: <http://www.opentext.com>. [Site visited on 27/03/2001].

PROFUSION. 2000. Home page. [Online]. Available: <http://www.profusion.com>. [Site visited on 27/06/2000].

SAVVYSEARCH. 2000. Home page. [Online]. Available: <http://www.savvysearch.com>. [Site visited on 19/06/2000].

SEARCH. 2000. Home page. [Online]. Available: <http://www.search.com>. [Site visited on 28/07/2000].

SNAP. 2000. Home page. [Online]. Available: <http://www.snap.com>. [Site visited on 14/06/2000].

WEBCRAWLER. 2000. Home page. [Online]. Available: <http://www.webcrawler.com>. [Site visited on 01/06/2000].

YAHOO! 2000. Home page. [Online]. Available: <http://www.yahoo.com>. [Site visited on 15/06/2000].

### **SEARCH ENGINE HOME PAGES: SLEEPING SERVERS**

(Sleeping Servers are search engine servers in countries, which have a local time of between 23h00 and 07h00 relative to the user - see Section 5.3.6).

#### **AUSTRALIA**

ALTAVISTA. 1999. Home page. [Online]. Available: <http://www.altavista.yellowpages.com.au/cgi-bin/telstra?pg=q>. [Site visited on 01/02/1999].

EXCITE. 1999. Home page. [Online]. Available: <http://au.excite.com/>. [Site visited on 25/01/1999].

YAHOO! 1999. Home page. [Online]. Available: <http://www.yahoo.com.au/> [Site visited on 15/02/1999].

#### **ASIA**

ALTAVISTA. 1999. Home page. [Online]. Available: <http://altavista.skali.com.my/>. [Site visited on 16/02/1999].

#### **CANADA**

ALTAVISTA. 1999. Home page. [Online]. Available: <http://www.altavistacanada.com/>. [Site visited on 21/02/1999].

YAHOO! 1999. Home page. [Online]. Available: <http://ca.yahoo.com/>. [Site visited on 09/02/1999].

**UNITED KINGDOM**

EXCITE. 1999. Home page. [Online]. Available: <http://www.excite.co.uk/>.  
[Site visited on 03/02/1999].

LYCOS. 1999. Home page. [Online]. Available: <http://www.lycos.co.uk/>.  
[Site visited on 17/02/1999].

YAHOO!. 1999. Home page. [Online]. Available: <http://www.yahoo.co.uk/>.  
[Site visited on 26/02/1999].

**SOUTH AFRICA**

ANANZI. 1999. Home page. [Online]. Available: <http://www.ananzi.co.za/>.  
[Site visited on 05/02/1999].

MAX. 1999. Home page. [Online]. Available: <http://www01.max.co.za/>.  
[Site visited on 14/02/1999].

ZEBRA. 1999. Home page. [Online]. Available: <http://www.zebra.co.za/>.  
[Site visited on 06/02/1999].

**OTHER REFERENCES**

ADAMASTOR. 2001. Origins & aims. [Online]. Available:  
<http://www.adamastor.co.za/Academic/Infolit/origins.htm> [Site visited on  
28/06/2001].

ALLAN, G. and SKINNER, C. 1991. Handbook for research students in the  
social sciences. London: Falmer.

ALRECK, P.L. and SETTLE, R.B. 1985. The research survey handbook.  
Homewood, IL: Irwin.

APPLEBEE, A., CLAYTON, P., PASCOE, C. and BRUCE, H. 2000. Australian academic use of the Internet: implications for university administrators. Internet Research: Electronic Networking Applications and Policy. 10(2): 141-149.

ASIS. 2001. Zator company. [Online]. Available: <http://www.asis.org/Features/Pioneers/zator.htm> [Site visited on 10/04/2001].

AUSTIN, D. 1984. PRECIS – a manual of concept analysis and subject indexing. London: Oxford University Press.

BAEZA-YATES, R. and RIBEIRO-NETO, B. 1999. Modern information retrieval. Harlow, Essex: Addison Wesley Longman.

BAR-ILAN, J. 2000. The web as an information source on informetrics? A content analysis. Journal of the American Society for Information Science. 51(5): 432-443.

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

BARRY, C.L. 1998. Document representations and clues to document relevance. Journal of the American Society for Information Science. 49(14): 1293-1303.

BASILI, C. 1995. Subject searching for information: what does it mean in today's Internet environment? The Electronic Library. 13(5): 459-465.

BASU, G. 1995. Using Internet for reference: myths vs. realities. Computers in Libraries. 15(2): 38-39.

BEHR, A.L. 1988. Empirical research methods for the human sciences. 2nd ed. Durban: Butterworth.

BILAL, D. 2001. Children's use of the Yahoo!igans! web search engine: II. Cognitive and physical behaviors on research tasks. Journal of the American Society for Information Science and Technology. 52(2): 118-136.

BOULTON, R. 2001. An introduction to information retrieval. [Online]. Available: [http://www.omsee.com/developer/docs/intro\\_ir.html](http://www.omsee.com/developer/docs/intro_ir.html) [Site visited on 10/04/2001].

BRANDT, D.S. 1996. Relevancy and searching the Internet. Computers in Libraries. 16(8): 35-39.

BRUCE, H. 1998. User satisfaction with information seeking on the Internet. Journal of the American Society for Information Science. 49(6): 541-556.

BRUCE, H. 1999. Perceptions of the Internet: what people think when they search the Internet for information. Internet Research: Electronic Networking Applications and Policy. 9(3): 187-199.

BUCKLAND, M. and GEY, F. 1994. The relationship between recall and precision. Journal of the American Society for Information Science. 45(1): 12-19.

BURGIN, R. 1992. Variations in relevance judgements and the evaluation of retrieval performance. Information Processing and Management. 28(5): 619-627.

BURTON, P.F. 1999. Information professionals and the world wide web. Online & CD-ROM Review. 23(2): 103-104.

BUSH, V. 1945. As we may think. [Online]. Available: <http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm> [Site visited on 22/06/2001].

BUTNER, R. 2001. Out of order: Hypertext's past, present and future.  
[Online]. Available: <http://www.zdnet.com/yil/content/mag/9611/hyper9611.html>  
[Site visited on 20/03/2001].

CERN. 2001. A CERN invention you are familiar with: the world wide web.  
[Online]. Available: <http://public.web.cern.ch/Public/ACHIEVEMENTS/web.html>  
[Site visited on 20/03/2001].

CHEN, H., HOUSTON, A.L., SEWELL, R.R. and SCHATZ, B.R. 1998. Internet browsing and searching: user evaluations of category map and concept space techniques. Journal of the American Society for Information Science. 49(7): 582-603.

CHUN, T.Y. 1999. World wide web robots: an overview. Online & CD-ROM Review. 23(3): 135-142.

CLARKE, S.J. and WILLETT, P. 1997. Estimating the recall performance of web search engines. Aslib Proceedings. 49(7): 184-189.

CLEVERDON, C.W., MILLS, J. and KEEN, E.M. 1966. Factors determining the performance of indexing systems. Vol 1. Design, Part 1. Cranfield: Aslib Cranfield Research Project.

COETZEE, H.S. 1999. The Dublin core initiative as a metadata standard for information retrieval on the Internet. Meta-info Bulletin. 8(4): 9-20.

COHEN, L.B. and JACOBSON, T.E. 1999. Choosing tools to search the web. The Teaching Professor. 13(9): 1-4.

COLLINS, S. 1997. Seek and ye shall find. .net. January: 50-66.

CONNELL, T.H. 1995. Subject searching in online catalogs: metaknowledge used by experienced users. Journal of the American Society for Information Science. 46(7): 506-518.

CONTE, R. 1996. Guiding lights. Internet World. May: 41-44.

COOKSON, P.S. 2000. Implications of Internet technologies for higher education: North American perspectives. The Journal of Open and Distance Learning. 15(1): 71-80.

COOPER, M.D. 2001. Usage patterns of a web-based library catalog. Journal of the American Society for Information Science and Technology. 52(2): 137-148.

CORCORAN, M., DAGAR, L. and STRATIGOS, A. 2000. The changing roles of information professionals. Online. March/April: 28-34.

CORRITORE, C.L., HICKMAN, B.L., GRANDGENETT, N. and HITCHCOCK, R. 1999. A comparison of two laboratory-based approaches for teaching introductory computer science. Innovations in Education and Training International. 36(4): 292-301.

COURTOIS, M.P. and BERRY, M.W. 1999. Results ranking in web search engines. Online: 39-46.

COVER, R. 2001. Electronic text corpus of Sumerian literature. [Online]. Available: <http://www.oasis-open.org/cover/etcs1.html> [Site visited on 10/04/2001].

CRONJE, J.C. and CLARKE, P.A. 1999. Teaching "teaching on the Internet" on the Internet. South African Journal of Higher Education. 13(1): 213-226.

CTP. 1999. Profiles of South African Technikon: 1999 Report. Pretoria: Committee of Technikon Principals.

DAVIS, B.H. 1996. Teacher of the future. Journal of the American Society for Information Science. 47(11): 849-853.

DE JAGER, K. 1995. What is the use of the library? A search for appropriate measures with which to evaluate the functions of an academic library in South Africa in the 1990s. Doctor of Philosophy thesis, University of Cape Town. (Unpublished).

DE JAGER, K. and SAYED, Y. 1998. Aspects of information literacy at five institutions of higher education in the Western Cape. South African Journal of Higher Education. 12(2): 197-203.

DEVLIN, B., and BURKE, M. 1997. Internet: the ultimate reference tool? Internet Research. 7(2): 101-108.

DILLMAN, D.A. 1978. Mail and telephone surveys – the total design method. New York, NY: John Wiley.

DONG, X. and SU, L.T. 1997. Search engines on the world wide web and information retrieval from the Internet: a review and evaluation. Online & CD-ROM Review. 21(2): 67-80.

DRAGAN, R.V. and BOSCARDIN, A. 1997. Taming the wild, wild web. PC Magazine. March 25: 192.

DUNN, R. 1996. Growth of the Internet. NFAIS Newsletter. April/May: 77-78.

DWYER, P. 1997. Search Engines – The Future. .net. July: 97-100.

EDLING, R.J. 2000. Information technology in the classroom: experiences and recommendations. Campus-Wide Information Systems. 17(1): 10-15.

ELLIS, D. 1996. Progress and problems in information retrieval. London: Library Association Publishing.

- KRANER, J. and WILLETT, P. 1996. On the creation of hypertext text documents: measurement of retrieval effectiveness. Journal of American Society for Information Science. 47(4): 287-300.
- EVANS, B. 1999. The authors of academic library home pages: their identity, training and dissemination of Web construction skills. Internet Research: Electronic Networking Applications and Policy. 9(4): 309-319.
- EVANS, P.B. and WURSTER, T.S. 1997. Strategy and the new economics of information. Harvard Business Review. September/October: 71-82.
- FARMER, J., RICHARDSON, A. and LAWTON, S. 1999. Improving access to information for nursing staff in remote areas: the potential of the Internet and other networked information resources. International Journal of Information Management. 19(1): 49-62.
- FELDMAN, S. 1998. Web search services in 1998: trends and challenges. [Online]. Available: <http://www.infotoday.com/searcher/jun98/story2.htm> [Site visited on 08/12/2000].
- FELDMAN, S. 1999. Natural language processing in information retrieval. Online. May/June: 62-72.
- FIDEL, R. 1982. Constructing a pattern model of man-machine interaction in online literature searching. [Online]. Available: <http://www.clis.umd.edu/research/dissertations/fidel.html> [Site visited on 22/02/2001].
- FIDEL, R. 1986. Towards expert systems for the selection of search keys. Journal of the American Society for Information Science. 37(1): 37-44.
- FIDEL, R. 1991a. Searchers' selection of search keys: I. The selection routine. Journal of the American Society for Information Science. 42(7): 490-500.

GARDNER, G. 1978. Social surveys for social planners. Milton Keynes: Open University.

GARFIELD, E. 2001. A tribute to Calvin N. Mooers, a pioneer of information retrieval. [Online]. Available: [http://the-scientist.com/yr1997/mar/comm\\_970317.html](http://the-scientist.com/yr1997/mar/comm_970317.html)  
[Site visited on 10/04/2001].

GARMAN, N. 1999. Meta search engines. Online. 23(3): 74-78.

GAUCH, S. and SMITH, J.B. 1993. An expert system for automatic query reformation. Journal of the American Society for Information Science. 44(3): 124-136.

GAWA, N. 2001. Arming ourselves for recognition of prior learning. [Online]. Available: <http://www.saardhe.ac.za/sajhe/99132p0.html#gawe>. [Site visited on 28/06/2001].

GILLIES, J., and CAILLIAU, R. 2000. How the web was born. Oxford: Oxford University Press.

GITTLEN, S. 1997. Revolutionising the 'net search engine. Computing S.A. 27 October: 17.

GOOD, C.V. 1963. Introduction to educational research. New York, NY: Appleton-Century-Crofts.

GREEK LIBRARIES. 2001. Libraries. [Online]. Available: [http://www.libraries.gr/en\\_messop.htm](http://www.libraries.gr/en_messop.htm). [Site visited on 11/05/2001].

GREEN, R. 1995. Topical relevance relationships: I. Why topic matching fails. Journal of the American Society for Information Science. 46(9): 646-653.

GREEN, R. and BEAN, C.A. 1995. Topical relevance relationships: II. An exploratory study and preliminary typology. Journal of the American Society for Information Science. 46(9): 654-662.

GREISDORF, H. 2001. Relevance: an interdisciplinary and information science perspective. [Online]. Available: <http://inform.nu/Articles/Vol3/v3n2p67-72.pdf>. [Site visited on 12/04/2001].

GRIBBLE, C. 2001. History of the web beginning at CERN. [Online]. Available: [http://www.hitmill.com/internet/web\\_history.html](http://www.hitmill.com/internet/web_history.html). [Site visited on 18/3/2001].

GRIMES, D.J. and BOENING, C.H. 2001. Worries with the web: a look at student use of web resources. College and Research Libraries. 62(1): 11-23.

GUPTA, A. and JAIN, R. 1997. Visual information retrieval. Communications of the ACM. 40(5): 71-79.

HABIB, D.P. and BALLIOT, R.L. 1999. How to search the world wide web: a tutorial for beginners and non-experts. [Online]. Available: <http://www.ultranet.com/~egrilib/totur.htm> [Site visited on 6/3/1999].

HARTER, S.P. 1996. Variations in relevance assessments and the measurement of retrieval effectiveness. Journal of the American Society for Information Science. 47(1): 37-49.

HARTER, S.P. and CHENG, Y.R. 1996. Colinked descriptors: improving vocabulary selection for end-user searching. Journal of the American Society for Information Science. 47(4): 311-325.

HERSH, W.H., PENTECOST, J. and HICKAM, D. 1996. A task-oriented approach to information retrieval evaluation. Journal of the American Society for Information Science. 47(1): 50-56.

HIRSH, S.G. 1999. Children's relevance criteria and information seeking on electronic resources. Journal of the American Society for Information Science 50(14): 1265-1283.

HOCK, R. 1999. Web search engines - features and commands. Online. May/June: 24-28.

HOFF, S.R.C. 1997. AltaVista vs. Excite vs. Hotbot vs. Infoseek: which is the one to rely on? [Online]. Available: <http://www4.zdnet.com/pccomp/srchoff/srchoff.html> [Site visited on 1/10/1997].

HOFFMAN, D., KALSBECK, W.D., NOVAK, T.P. 2001. Internet use in the United States: 1995 baseline estimates and preliminary market segments. [Online]. Available: <http://www2000.ogsm.vanderbilt.edu/baseline/1995.Internet.estimate.html> [Site visited on 17/03/2001].

HOINVILLE, G. and JOWELL, R. and associates. 1982. Survey research practice. London: Heinemann Educational.

HOLLAND, M.P. and POWELL, C.K. 1995. A longitudinal survey of the information seeking and use habits of some engineers. College & Research Libraries. 56(1): 7-15.

HOWCROFT, D. and MITEV, N. 2000. An empirical study of Internet usage and difficulties among medical practice management in the UK. Internet Research: Electronic Networking Applications and Policy. 10(2): 170-181.

HSIEH-YEE, I. 1993. Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. Journal of the American Society for Information Science. 44(3): 161-174.

HUWE, T.K. 1999. New search tools for multidisciplinary digital libraries. Online. March/April: 67-74.

- HUYSAMEN, G.K. 1994. Methodology for the social and behavioural sciences. Midrand: Southern.
- INGWERSEN, P. 1992. Information retrieval interaction. London: Taylor Graham.
- IVES, B. and JARVENPAA, S.L. 1996. Will the Internet revolutionize business education and research? Sloan Management Review. Spring: 33-41.
- JACSO, P. 1999a. Database selection tools. Online & CD-ROM Review. 23(4): 227-229.
- JACSO, P. 1999b. Savvy searchers do ask for directions. Online & CD-ROM Review. 23(2): 99-102.
- JACKSON, E.B. 1971. The engineer as reluctant information user – a remedial plan. Proceedings of the International Conference on Training for Information Work. Rome. November 15-19: 430-434.
- JANES, J. 2000. Sampling: you don't have to ask everybody. Library Hi Tech. 18(1): 97-101.
- JANSEN, B.J., and POOCH, U. 2001. A review of web searching studies and a framework for future research. Journal of the American Society for Information Science and Technology. 52(3): 235-246.
- JONES, S. 1996. Indexing the Internet – a job for machine? (or does it take human intervention?). NFAIS Newsletter. April/May: 78-80.
- KASSLER, H.S. 1997. Mining the Internet for competitive intelligence. Online. September/October: 34-45.
- KEENAN, S. 1997. Checking on the arrows. Online & CD-ROM Review. 21(4): 236-238.

KHAN, K., and LOCATIS, C. 1998a. Searching through cyberspace: the effects of link cues and correspondence on information retrieval from hypertext on the world wide web. Journal of the American Society for Information Science. 49(14): 1248-1253.

KHAN, K., and LOCATIS, C. 1998b. Searching through cyberspace: the effects of link display and link density on information retrieval from hypertext on the world wide web. Journal of the American Society for Information Science. 49(2): 176-182.

KILLEN, R. 1997. Outcomes based education: rethinking teaching. Ekonomie. 10(1): 25-32.

KOTHARI, C.R. 1991. Research methodology methods & techniques. New Delhi: Wiley Eastern.

KRISTULA, D. 2001. The history of the Internet. [Online]. Available: <http://www.davesite.com/webstation/net-history.shtml>. [Site visited: 18/03/2001].

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.

KWMC. 2001. The Internet: what it is. [Online]. Available: <http://www.kwmc.on.ca/resources/internet/history.html>. [Site visited: 14/03/2001].

LABOVITZ, S. and HAGEDORN, R. 1981. Introduction to social research. New York, NY: McGraw-Hill.

LANCASTER, F.W. 1978. Information retrieval systems: characteristics, testing and evaluation. New York, NY: John Wiley.

LANDONI, M. and BELL, S. 2000. Information retrieval techniques for evaluating search engines: a critical overview. Aslib Proceedings. 52(3): 124-129.

LARGE, A and BEHESHTI, J. 2000. The web as a classroom resource: reactions from the users. Journal of the American Society for Information Science. 51(12): 1069-1080.

LARGE, A., TEDD, L.A. and HARTLEY, R.J. 1999. Information seeking in the online age: principles and practice. London: Bowker-Saur.

LAWRENCE, S. and GILES, L. 1999. Accessibility of information on the web. Nature. 400: 107-109.

LEEDY, P.D. 1985. Practical research – planning and design. New York, NY: Macmillan.

LEIGHTON, H.V. and SRIVASTAVA, J. 1999. First 20 precision among world wide web search services (search engines). Journal of the American Society for Information Science. 50(10): 870-881.

LESSER, L., HORLING, R., KLASSNER, F., RAJA, A., WAGNER, T. and ZHANG S.X.Q. 2000. BIG: an agent for resource-bounded information gathering and decision making. Artificial Intelligence. 118: 197-244.

LEWIS, D.A. 1980. Today's challenge-tomorrow's choice: change or be changed or the doomsday scenario mk2. Journal of Information Science. 2: 59-74.

LIDSKY, D. and KWON, R. 1997. Searching the Net. PC Magazine. December 2: 227-255.

LOSEE, M. and PARIS, L.A.H. 1999. Measuring search-engine quality and query difficulty: ranking with Target and Freestyle. Journal of the American Society for Information Science. 50(10): 882-889.

LOVELL, K. and LAWSON, K.S. 1970. Understanding educational research. London: University of London Press.

LYNCH, C.A. 2001. When documents deceive: trust and provenance as new factors for information retrieval in a tangled web. Journal of the American Society for Information Science and Technology. 52(1): 12-17.

MAJID, S. and ABAZOVA, A.F. 1999. Computer literacy and use of electronic information sources by academics: a case study of International Islamic University Malaysia. Asian Libraries: The Library and Information Management Journal for Asia and the Pacific. 8: 70-81.

MALAN, S.P.T. 1999. Imperatives for technology-enhanced delivery systems for providing education in South Africa. South African Journal for Higher Education. 13(1): 59-64.

MANNO, B.V. 1994. Outcome-based education: has it become more affliction than cure? [Online]. Available: <http://www.amexp.org/publications/education/edu5.htm>. [Site visited: 06/08/2000].

MARAIS, P.J.J.G. and MARAIS, A.F. 1999. Science education for the new millennium: an exciting challenge. South African Journal of Higher Education. 13(2): 82-89.

MARCELLA, R., BAXTER, G. 1999. The information needs and the information seeking behaviour of a national sample of the population in the United Kingdom, with special reference to needs related to citizenship. Journal of Documentation. 55(2): 159-183.

MARCHIONINI, G. 1995. Information seeking in electronic environments. New York, NY: Cambridge University Press.

MARTIN, W.A. 1982. Helping the less experienced user. Proceedings of the 6<sup>th</sup> International Online Information Meeting. London. 67-76.

MASON, E.J. and BRAMBLE, W.J. 1989. Understanding and conducting research. New York, NY: McGraw-Hill.

MCPL. 2001. A brief history of the Internet – statistics from the third MIDS Internet demographics survey. [Online]. Available: <http://www.monroe.lib.in.us/~net101/hist.html>. [Site visited on 15/03/2001].

MELLON, C.A. 1999. Technology and the great pendulum of education. Journal of Research on Computing in Education. 32(1): 28-35.

MICKEY, B. 1999. A web search trifecta. Online. May/June: 79-82.

MICROSOFT. 2000. Republic of South Africa. Microsoft Encarta Encyclopedia 2000, World English Edition. [CD-ROM].

MIZZARO, S. 1997. Relevance: the whole history. Journal of the American Society for Information Science. 48(9): 810-832.

MIZZARO, S. 1998. How many relevances in information retrieval? Interacting with Computers. 16: 303-320.

MOHSIN, S.M. 1984. Research methods in behavioural sciences. Calcutta: Orient Longman.

MORRISON, P. 2000. Online privacy and the home user – an investigation of the online environment encountered by the home user, with suggestions on how to increase personal privacy. B. Tech Degree (Information Technology), Cape Technikon. (Unpublished).

MOUTON, J. 1996. Understanding social research. Pretoria: J.L. van Schaik Academic.

NAPTOSA. 2000. Understanding SAQA and the NQF. NAPTOSA Newsletter. January: 1-8.

NCHE. 2000. Apartheid's higher education legacy. [Online]. Available: <http://www.star.hsrc.ac.za/nche/final/tg1/tgcs6.html>. [Site visited on 1/08/2000].

NICHOLSON, S. 2000. Raising reliability of web search tool research through replication and chaos theory. Journal of the American Society for Information Science. 51(8): 724-729.

NIGOHOSIAN, R.H. 1997. Scholarly Internet research: is it real? [Online]. Available: <http://www.slcc.edu/b10/ecn274/ethic.html>. [Site visited on 16/10/1997].

NOBLES, T. 1997. What will it be, directory or search engine? [Online]. Available: <http://www.phoenixmlm.com/sengine.html>. [Site visited on 16/10/1997].

NOTESS, G.R. 1995a. The Infoseek databases. Online. August/September: 85-87.

NOTESS, G.R. 1995b. Searching the world wide web: Lycos, Webcrawler and more. Online. July/August: 48-53.

NOTESS, G.R. 1997. Internet search techniques and strategies. Online. July/August: 63-66.

NOTESS, G.R. 1999a. Internet search engine update. Online. March/April: 17.

NOTESS, G.R. 1999b. On-the-fly search engine analysis. Online.  
September/October: 63-66.

NOTESS, G.R. 1999c. Rising relevance in search engines. Online.  
May/June: 84-86.

NUA. 2001. How many online? [Online]. Available:  
[http://www.nua.ie/surveys/how\\_many\\_online/index.html](http://www.nua.ie/surveys/how_many_online/index.html). [Site visited on  
10/04/2001].

O'LEARY, M. 2000. New roles come of age. Online. 24(2): 21-25.

OLIVER, R. and OLIVER, H. 1997. Using context to promote learning from  
information-seeking tasks. Journal of the American Society for Information  
Science. 48(6): 519-526.

OPPENHEIM, A.N. 1992. Questionnaire design, interviewing and attitude  
measurement. New ed. London: Pinter.

OPPENHEIM, C., MORRIS, A., McKNIGHT, C. and LOWLEY, S. 2000. The  
evaluation of www search engines. Journal of Documentation. 56(2): 190-211.

PACK, T. and PEMBERTON, J. 1999a. The cutting-edge library at the Los  
Alamos national laboratory. Online. March/April: 34-42.

PACK, T. and PEMBERTON, J. 1999b. The cutting-edge library at Shell  
research. Online. July/August: 28-33.

PAGE, A. 1997. The search is over. [Online]. Available:  
<http://www4.zdnet.com/pccomp/features/fea1096/sub2.html>. [Site visited on  
16/10/1997].

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.

PEMBERTON, J. and PACK, T. 1999. The cutting-edge library at Hewlett-Packard. Online. September/October: 30-36.

PERRY, T.T., PERRY, L.A. and HOSACK-CURLIN, K. 1998. Internet use by university students: an interdisciplinary study on three campuses. Internet Research: Electronic Networking Applications and Policy. 8(2): 136-141.

PETERSEN, R.G. 1985. Design and analysis of experiments. New York, NY: Marcel Dekker.

POTENZA, E. and JANSEN, J. 1998. Outcomes-base education: will it benefit our children? Fairlady. January 21: 52-56.

QUILLING, R.D., ERWIN, G.J. and PETKOVA, O. 1999. Active learning: issues and challenges for information systems and technology. South African Computer Journal. 24: 5-14.

RADCLIFF, C., DU MONT, M. and GATTEN, J. 1993. Internet and reference services: implications for academic libraries. Library Review. 42(1): 15-19.

ROBBINS, D. 1999. Tertiary education: the size and shape of things to come. [Online]. Available: <http://www.mg.co.za/mg/news/99nov2/24nov-education.html>. [Site visited on 09/08/2000].

ROBERTSON, S.E. 1977. The probabilistic character of relevance. Information Processing & Management. 13: 247-251.

ROBERTSON, S.E. 2001. Conflicting philosophies. [Online]. Available: <http://www.soi.city.ac.uk/research/cisr/ser/ucla/node3.html> [Site visited on 10/04/2001].

ROSS, N.C.M. and WOLFRAM, D. 2000. End user searching on the Internet: an analysis of term pair topics submitted to the Excite search engine. Journal of the American Society for Information Science. 51(10): 949-958.

SAPSFORD, R. and JUPP, V. 1996. Data collection and analysis. London: Sage.

SARACEVIC, T. 1975. Relevance: a review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science. 26: 321-343.

SARACEVIC, T. 1999. Information science. Journal of the American Society for Information Science. 50(12): 1051-1063.

SARACEVIC, T. and KANTOR, P. 1988a. A study of information seeking and retrieving. II. Questions and effectiveness. Journal of the American Society for Information Science. 39(3): 177-196.

SARACEVIC, T. and KANTOR, P. 1988b. A study of information seeking and retrieving. III. Searchers, searches and overlap. Journal of the American Society for Information Science. 39(3): 197-216.

SARACEVIC, T., BRAGA, G.M. and AFOLAYAN, M.A. 1984. Issues in information science education in developing countries. Journal of the American Society for Information Science. 36(3): 192-199.

SARACEVIC, T., KANTOR, P., CHAMIS, A.Y. and TRIVISON, D. 1988. A study of information seeking and retrieving. I. Background and methodology. Journal of the American Society for Information Science. 39(3): 161-176.

SCHAMBER, L. 1994. Relevance and information behavior. Annual Review of Information Science and Technology. 29: 3-48.

SCHWARTZ, C. 1998. Web search engines. Journal of the American Society for Information Science. 49(11): 973-982.

SELWYN, N. 1999. Educational superhighways – in the public or private interest? Internet Research: Electronic Networking Applications and Policy. 9(3): 225-231.

SHA, V.T. 1995. Cataloguing Internet resources: the library approach. The Electronic Library. October. 13(5): 467-476.

SHERMAN, C. 1999. The future of web search. Online. May/June: 54-61.

SHERMAN, C. 2000. Reference resources on the web. Online. January/February: 52-56.

SIEGFRIED, S., BATES, M.J. and WILDE, D.N. 1993. A profile of end-user searching behavior by humanities scholars: the Getty online searching project report no. 2. Journal of the American Society for Information Science. 44(5): 273-291.

SINGH, A. and LIDSKY, D. 1996. All-out search. PC Magazine. December: 213-245.

SMIT, G.J. 1983. Navorsingsmetodes in die gedragswetenskappe. Pretoria: Haum.

SNYDER, H. and ROSENBAUM, H. 1999. Can search engines be used as tools for web-link analysis? A critical view. Journal of Documentation. 55(4): 375-384.

SPADY, W.G. 1994. Outcome-based education: critical issues and answers. Virginia: American Association of School Administrators.

SPINK, A. 1996. Multiple search sessions model of end-user behavior: an exploratory study. Journal of the American Society for Information Science. 47(8): 603-609.

SPINK, A. and GREISDORF, H. 2001. Regions and levels: measuring and mapping users' relevance judgements. Journal of the American Society for Information Science and Technology. 52(2): 161-173.

SPINK, A. and LOSEE, R.M. 1996. Feedback in Information retrieval. Annual Review of Information Science and Technology. 31: 33-78.

SPINK, A. and SARACEVIC, T. 1997. Interaction in information retrieval: selection and effectiveness of search terms. Journal of the American Society for Information Science. 48(8): 741-761.

SPINK, A., BATEMAN, J. and JANSEN, B.J. 1999. Searching the web: a survey of Excite users. Internet Research: Electronic Networking Applications and Policy. 9(2): 117-128.

SPINK, A. and others. 2001. Searching the web: the public and their queries. Journal of the American Society for Information Science and Technology. 52(3): 226-234.

STRATEGIS. 2000. Internet review. [Online]. Available: <http://strategis.comnow.com/allpubs.lasso?-database=reports.fp3&-layout=ma>. [Site visited on 14/4/2000].

STUBENRAUCH, V.R., VICKERY, B. and RUPPERT, H.A. 1999. The digital library as a personal knowledge platform: Liberation. [Online]. Available: <http://www.b-l-t-online.de/archiv/199-01/fachbeitraege/beitrag02/02.htm> [Site visited on 15/12/2000].

- SU, L.T. 1993. Is relevance an adequate criterion for retrieval system evaluation: an empirical inquiry into the user's evaluation. Proceedings of the 56th Annual Meeting of the American Society for Information Science. 24-28 October. Columbus, OH: 93-103.
- SU, L.T. 1994. The relevance of recall and precision in user evaluation. Journal of the American Society for Information Science. 45(3): 207-217.
- SULLIVAN, D. 1999a. Crawling under the hood. Online. May/June: 30-38.
- SULLIVAN, D. 1999b. Search Engine Report. Available: e-mail message [received on 05/01/1999].
- SULLIVAN, D. 2000. Search engine sizes. [Online]. Available: <http://www.searchenginewatch.com>. [Site visited on 22/7/2000].
- SUTCLIFFE, A.G., ENNIS, M. and WATKINSON, S.J. 2000. Empirical studies of end-user information searching. Journal of the American Society for Information Science. 51(13): 1211-1231.
- SUTTON, S.A. 1999. Conceptual design and development of a metadata framework for educational resources on the Internet. Journal of the American Society for Information Science. 50(13): 1182-1192.
- TATE, M. and ALEXANDER, J. 1996. Teaching critical evaluation skills for world wide web resources. Computers in Libraries. 6: 49-55.
- TAUBES, G. 1995. Indexing the Internet. Science. 269: 1354-1356.
- TEARLE, P., DILLON, P. and DAVIS, N. 1999. Use of information technology by English university teachers. Developments and trends at the time of the National Inquiry into Higher Education. Journal of Further and Higher Education. 23(1): 5-15.

TEGENBOS, J. and NIEUWENHUYSEN, P. 1997. My kingdom for an agent? Evaluation of Autonomy, an intelligent search agent for the Internet. Online & CD-ROM Review. 21(3): 139-148.

THOMAS, D.H. 2001. The effect of interface design on item selection in an online catalog. Library Resources & Technical Services. 45(1): 20-46.

TONTA, Y.A. 2001. Failure analysis in document retrieval systems: a critical review of studies. [Online]. Available:

<http://yunus.hacettepe.edu.tr/~tonta/yayinlar/phd/bolum-3.htm> [Site visited on 18/04/2001].

TRAC, 2001. Information retrieval & Zatocoding. [Online]. Available:

<http://tracfoundation.org/mooers/infosci/infosci.htm> [Site visited on 10/04/2001].

TSICHRITZIS, D. 1997. How to surf the technology waves we created.

Communications of the ACM. 40(2): 49-54.

VAN DER MERWE, J. 2000. Adamastor Trust Publishing Liaison Office:

copyright and course notes. Cape Town: Adamastor Trust.

VAN ZIJL, C. and GERICKE, E.M. 1998. Information-seeking patterns of

artists and art scholars at the Vaal Triangle Technikon. South African Journal of Library and Information Science. 66(1): 23-33.

VENDITTO, G. 1996. Search engine showdown. Internet World. May: 79-80.

VICKERY, B. 1997. Knowledge discovery from databases: an introductory

review. Journal of Documentation. 53(2): 107-122.

VOORBIJ, H.J. 1999. Searching scientific information on the Internet: a Dutch

academic user survey. Journal of the American Society for Information Science. 50(7): 598-615.

VOSS, N.D. 2001. Jones telecommunications & multimedia encyclopedia: George Boole overview. [Online]. Available: <http://www.digitalcentury.com/encyclo/update/boole.html>. [Site visited on 24/03/2001].

VROOM, V.H. 1964. Work and motivation. New York, NY: John Wiley.

WALLACE, R.M., KUPPERMAN, J. and KRAJCIK, J. 2000. Science on the web: students online in a sixth-grade classroom. The Journal of the Learning Sciences. 9(1): 75-104.

WEIDEMAN, M. 1994. A critical evaluation of the destructive impact of computer viruses on files stored by personal computer users. Master's Diploma in Technology (Information Technology), Cape Technikon. (Unpublished).

WEIDEMAN, M. 2000a. Help menu. [Online]. Available: <http://www.mwe.co.za/seamodhelp.htm>. [Site visited on 13/6/2000].

WEIDEMAN, M. 2000b. Search engines. [Online]. Available: <http://www.mwe.co.za/seaengin.htm>. [Site visited on 23/6/2000].

WHITE, H.S. 1995. At the crossroads: librarians on the information superhighway. Englewood, CO: Libraries Unlimited.

WILLIAMS, J. 1996. Bots and other internet beasts. Indianapolis, IN: Sams.

WINSHIP, I.R. 1995. World-wide web searching tools: an evaluation. Vine. 99: 49-54.

WOLFRAM, D. and DIMITROFF, A. 1997. Preliminary findings on searcher performance and perceptions of performance in a hypertext bibliographic retrieval system. Journal of the American Society for Information Science. 48(12): 1142-1145.

WONG, P. 1998. Leapfrogging across the millenium: information technology in Singapore schools. Journal of Global Information Management. 6(1): 5-13.

WYLE, M.F. 2001. Information overload. [Online]. Available: <http://eiger.wyle.org/~mfw/diss/node12.html>. [Site visited on 10/04/2001].

YUAN, W. 1997. End-user searching behaviour in information retrieval: a longitudinal study. Journal of the American Society for Information Science. 48(3): 218-234.

University of Cape Town

# GLOSSARY

University of Cape Town

**Bandwidth**

The data carrying capacity of an electronic communications channel.

**Boolean search**

A search allowing the inclusion or exclusion of documents containing certain words through the use of operators such as AND, NOT, OR and NEAR – also called a Logical search.

**Bot**

A computer program designed to travel across the Internet automatically, gathering information about websites in the process. Also called robot, spider or crawler.

**Browser program**

A program run by an Internet user that can display world wide web pages. The first browser program was Mosaic, and current commonly used programs include Microsoft Internet Explorer and Netscape Navigator.

**Cache memory**

High speed, high cost and therefore low capacity random access memory normally used to store often used information. Normally incorporated to speed up accessing of data by preventing slow accessing of disks.

**Clipboard**

A feature of the Windows operating system that allows the transfer of data from one application to another.

**Crawler**

A computer program designed to travel across the Internet automatically, gathering information about websites in the process. Also called robot, bot or spider.



**Information Technology**

A field of study which concentrates on the use of computer technology to enhance man's learning and understanding.

**Internet**

A worldwide collection of connected computers.

**Intranet**

A network in a company which could contain parts of the Internet, and limits access by outside users.

**Keyword search**

A search for documents containing one or more words that are specified by a user.

**Learner**

A person who is involved in studying a certain topic with the aim of understanding it and putting it to use.

**Lecture**

A session where an educator communicates with learners, mostly in a classroom situation, in an attempt to divulge knowledge, skills or perceptions.

**Logical search**

A search allowing the inclusion or exclusion of documents containing certain words through the use of operators such as AND, NOT, OR and NEAR – also called a Boolean search.

**Meta search engine**

A program which appears to be a normal search engine to the user, but in fact contains no index of its own. It uses the index of one or more other search engines (in parallel, if more than one) to extract information.



**Precision**

The degree to which a search engine lists documents matching a query. The more matching documents that are listed, the higher the precision.

**Query**

A word, number, phrase(s), operator(s) or sentence(s) which expresses the user's information need in a language which the search engine can understand.

**Ranking**

A method used by a search engine to sort and display search results in such a way that the most relevant answer appears first, the second most relevant one second, etc.

**Recall**

Related to precision, this is the degree to which a search engine returns all the matching documents in a collection. There may be 100 matching documents, but a search engine may only find 80 of them. It would then list these 80 and have a recall of 80%.

**Relevance**

How well a retrieval system provides documents which contain the information a user is looking for, as measured by the user.

**Robot**

A computer program designed to travel across the Internet automatically, gathering information about websites in the process. Also called spider, bot or crawler.

**Search engine**

A program which allows a user to specify a query, and then attempts to find information in its index file.

**Searching**

The process of attempting to find useful information within a large base of unordered data.

**Server**

A computer which stores programs and/or data which is requested and used by computer users from other computers, and which manages these requests in an orderly and timely fashion.

**Spamming**

The process of flooding a computer or user with (normally) useless information, sometimes in the form of e-mail messages, in an attempt to de-activate this computer or make productive work by the user impossible.

**Spider**

A computer program designed to travel across the Internet automatically, gathering information about websites in the process. Also called robot, bot or crawler.

**Stemming**

The ability to include the "stem" of words in a search. For example, stemming allows a user to enter "swimming" and obtain results also for the stem word "swim". Also called internal truncation or interfixing, which enables matching when variant characters appear (for example colour and color).

**Stop words**

Conjunctions, prepositions, articles and other words such as THE, AND, TO and A that often appear in documents yet alone may contain little meaning. Depending on the environment, any word can be nominated as a stop word. When searching within a collection of documents which covers assembler programming, for example, the (otherwise meaningful) word "assembler" could be a stop word.

**Study Aid**

A means that will increase the speed and ease of learning for a learner.

**Tertiary level**

The level of study which follows the last year of schooling.

**Traditional resources**

Textbooks, libraries, notes, videos, computer programs.

**Transparency**

A see-through page containing writing, drawings or any other study material which can be projected on a wall or screen by using an overhead projector.

**URL**

Universal Resource Locator – the standard name for the address of a website on the Internet.

**Windows**

A commonly used personal computer operating system whose main feature is the ability to execute different programs in separate frames, called windows, on the screen.