

Title	A task-based scientific paper recommender system for literature review and manuscript preparation(Dissertation)
Author(s)	Sesagiri Raamkumar, Aravind
Citation	Aravind, S. R. (2018). A task-based scientific paper recommender system for literature review and manuscript preparation. Doctoral thesis, Nanyang Technological University, Singapore.
Date	2018-10-08
URL	http://hdl.handle.net/10220/46243
Rights	

**A TASK-BASED SCIENTIFIC PAPER RECOMMENDER
SYSTEM FOR LITERATURE REVIEW AND
MANUSCRIPT PREPARATION**

SESAGIRI RAAMKUMAR ARAVIND

WEE KIM WEE SCHOOL OF COMMUNICATION AND INFORMATION

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2018

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors Professor Schubert Foo and Dr. Natalie Pang, for their supervision and support throughout my study. This thesis was largely possible only because of their guidance and assistance.

Prof Foo's comments have always been very insightful and his remarks helped me to identify issues specific to my research at different points of my candidature. Prof Foo also made me realize the importance of working on research problems that were worth pursuing. Dr. Pang's thoughtful suggestions have helped me to see the bigger picture of my work, with more emphasis on academic style thinking. Dr. Pang also helped me in writing my first set of journal papers. With the diverse perspectives of my supervisors on different topics, I could fine tune my research.

I would like to thank my internal and external thesis examiners as well as oral exam examiners for their generous contributions and suggestions, which have helped to improve my thesis in various aspects. Also thanks to all professors in the school for sharing their invaluable knowledge and providing kind support.

I would like to take this opportunity to express my appreciation to all staff in the school for being helpful and supportive. My appreciation also goes to my officemates, who have provided both intellectual and emotional support over the last five years.

I thank my beloved parents, my younger brother, my elder cousin Lakshminarayanan, cousin-in-law Jayanthi and my good friends Gopal, Senthil, Muthu and Divya for their continuous support and encouragement.

ABSTRACT

In the domain of scholarly communication lifecycle, recommender systems have been built to provide research papers for researchers' explicit and implicit information needs. Previous studies (Jardine, 2014; Mcnee, 2006) have employed an algorithmic approach of providing solutions to researcher's tasks. The characteristics of the tasks, their inter-relationships and intra-relationships with algorithms have been largely ignored since the focus has mainly been to propose different recommendation techniques on top of a variety of algorithms. Driven by these research gaps, the overarching goal of this research is to build an assistive system for helping researchers in finding papers for key Literature Review (LR) and Manuscript Preparatory (MP) tasks. To achieve this goal, two research objectives are proposed. The first objective is to identify an appropriate method to map the identified LR and MP tasks to relevant algorithms. The deliverable for this objective is a prototype assistive system that provides recommendations for three tasks. The second research objective is to evaluate whether the performance of the proposed recommendation techniques and the overall system are at the expected level. To address the research objectives, the research is divided into two interrelated studies.

In Study I, a university-wide survey was conducted on the topic of Inadequate and Omitted Citations in manuscripts (IOC). The 207 survey respondents were classified into manuscript reviewer and author groups. Survey results indicated that manuscript authors frequently miss citing seminal and topically-similar papers in journal manuscripts. The lack of experience in a specific research area was perceived as a major reason for IOC, followed by lack of overall research experience and the scenario of working in interdisciplinary research projects. Authors frequently needed

external assistance in finding interdisciplinary and topically-similar papers for LR. Based on the findings, two LR search tasks of building reading list and finding topically similar papers were shortlisted. A third task meant to help researchers in identifying unique and important papers from their final reading list was selected, thereby making it a total of three tasks for the assistive system. A prototype called Rec4LRW (Recommendations for Literature Review and Writing) system was developed for providing recommendations for the shortlisted three tasks. The system development was guided by a threefold intervention framework comprising of (i) task redesign for addressing the algorithmic improvements, (ii) task interconnectivity addressing the management of papers between the tasks and (iii) informational display features in the system's user-interface for expediting researcher's relevance judgment decisions.

The second research objective is addressed in Study II. As a part of Study II, an offline evaluation experiment and a user evaluation study were conducted. An extract of papers from the ACM Digital Library was used as the corpus for the evaluations. A total of 119 researchers who had experience in authoring research papers, participated in the user study. Predictors and correlates for the output quality measures were identified for each task. This study established the effectiveness of the three interventions in providing relevant recommendations. Graduate students and novice researchers found the recommendations and the overall system to be more useful and effective.

TABLE OF CONTENTS

CHAPTER ONE - INTRODUCTION	1
1.1 Background	1
1.2 Motivation.....	4
1.3 Research Objectives	5
1.4 Overview of Research Design	8
1.5 Assumptions.....	11
1.6 Dissertation Organizational Overview	12
CHAPTER TWO - LITERATURE REVIEW.....	14
2.1 Introduction.....	14
2.2 Part One - Scientific Information Seeking.....	14
2.2.1 IS Models Overview	15
2.2.2 Scientific IS Models.....	16
2.2.3 Scientific IS Behavior Studies	20
2.2.4 Summary of Part One.....	23
2.3 Part Two - Scientific Paper Recommendation Systems	26
2.3.1 Introduction.....	26
2.3.2 Classification of RS	27
2.3.3 Usage of RS in Scientific Research Lifecycle	28
2.3.4 Summary of Part Two	46
CHAPTER THREE - STUDY I: SURVEY ON INADEQUATE AND OMITTED CITATIONS (IOC) IN MANUSCRIPTS	48
3.1 Introduction.....	48
3.2 Aims of the Study	48
3.3 Method	50
3.3.1 Online Survey Instrument	50
3.3.2 Sampling	53
3.3.3 Participants.....	53
3.3.4 Data Analysis Procedures	56
3.4 Results.....	57
3.4.1 Instances of IOC	57

3.4.2	Reasons for IOC.....	59
3.4.3	Effects of IOC.....	60
3.4.4	Need for External Assistance in Finding Papers.....	61
3.4.5	Usage of Academic Information Sources	62
3.5	Discussion.....	65
3.6	Summary.....	70
CHAPTER FOUR - DESIGN AND DEVELOPMENT OF REC4LRW SYSTEM		
.....		74
4.1	Introduction.....	74
4.2	Identification of Base Features	74
4.2.1	References Count (RC).....	75
4.2.2	Grey Literature Percentage (GLP)	75
4.2.3	Coverage (C).....	76
4.2.4	Recency (RE).....	80
4.2.5	Textual Similarity (TS).....	80
4.2.6	Specificity (S)	81
4.2.7	Citation Count (CC).....	82
4.3	Recommendation Techniques for the Three Tasks.....	83
4.3.1	Ascertaining the Characteristics of the Three Tasks.....	84
4.3.2	Task Redesign.....	86
4.3.3	Task 1 - Building an Initial Reading List of Research Papers	88
4.3.4	Task 2 - Finding Similar Papers based on Set of Papers	94
4.3.5	Task 3 - Shortlisting Articles from Reading List for Inclusion in Manuscript	101
4.4	Rec4LRW System Design	105
4.4.1	Technical Details	106
4.4.2	Task Interconnectivity Mechanisms	107
4.4.3	User-Interface for the Three Tasks	108
4.4.4	Informational Display Features in the Rec4LRW User-Interface	111
4.5	Summary.....	117
CHAPTER FIVE - STUDY II: EVALUATION OF RECOMMENDATION TECHNIQUES AND REC4LRW SYSTEM.....		118

5.1	Introduction.....	118
5.2	Dataset used for Study II.....	118
5.3	Pre-study: Offline Evaluation of AKR Technique.....	120
5.3.1	Evaluated Techniques	120
5.3.2	Experiment Setup.....	122
5.3.3	Evaluation Metrics	123
5.3.4	Findings.....	124
5.4	User Evaluation Study	129
5.4.1	Objectives	129
5.4.2	Participant Recruitment and Screening.....	131
5.4.3	Participant Demographics	132
5.4.4	Study Procedure	137
5.4.5	Quantitative Data Analysis	139
5.4.6	Qualitative Data Analysis	186
5.5	Discussion	199
5.5.1	Influence of Task Redesign Intervention	199
5.5.2	Influence of Task Interconnectivity Intervention	202
5.5.3	Influence of Information Display Features Intervention	203
5.6	Summary.....	203
CHAPTER SIX - CONCLUSION.....		206
6.1	Introduction.....	206
6.2	Review of PhD Research	206
6.3	Contributions.....	211
6.4	Implications.....	214
6.4.1	Implications for Research	214
6.4.2	Implications for Practice	216
6.5	Limitations	217
6.6	Recommendations for Future Work.....	220
REFERENCES.....		224
APPENDIX A: IRB APPROVAL DOCUMENT.....		262
APPENDIX B: STUDY I DETAILS		264
APPENDIX C: PERFORMANCE TWEAKS FOR REC4LRW SYSTEM.....		273

APPENDIX D: CASE STUDY FOR GREY LITERATURE BOOSTING TECHNIQUE	277
APPENDIX E: CASE STUDY WITH THE COVERAGE FEATURES	282
APPENDIX F: FACETED CLASSIFICATION SCHEME FOR A TASK FROM LI & BELKIN (2008)	292
APPENDIX G: LIST OF 186 TOPICS USED FOR OFFLINE EVALUATION EXPERIMENT IN STUDY II	296
APPENDIX H: STUDY II DETAILS	303
APPENDIX I: CATEGORIES FROM THE QUALITATIVE ANALYSIS OF PARTICIPANT FEEDBACK FROM STUDY II.....	322
APPENDIX J: LIST OF PUBLICATIONS	328

LIST OF TABLES

Table 2.1 IS activities and problems identified in Chowdhury et al. (2014)	19
Table 2.2 Categorization schemes from previous studies of SPRS	28
Table 2.3 Categorization scheme for application for RS in research lifecycle tasks..	29
Table 3.1 Sections and the Corresponding Segment in the Online Questionnaire	51
Table 3.2 Experience of Participants on the Instances of IOC	58
Table 3.3 Mean Experience of Author Writing Groups for the Instances of IOC	59
Table 3.4 Participants Opinion on Reason for Inadequate and omitted citations (n = 146)	60
Table 3.5 Reviewers' Opinion on Effects of Missing Citations (n = 146)	60
Table 3.6 Authors' Experience on Effects of Inadequate and Omitted Citations (n = 207)	61
Table 3.7 Scenarios for External Assistance during Literature Review & Academic Writing (n = 207)	61
Table 3.8 Usage of Academic Information Sources and Related Papers Feature	63
Table 3.9 Usage of Scientific Paper Recommendation Services	64
Table 4.1 Facet Values for the Three Tasks Identified for the Assistive System	84
Table 4.2 Task redesign components	87
Table 4.3 Technologies and libraries used in Rec4LRW development	107
Table 4.4 Informational display features in Rec4LRW task screens	111
Table 4.5 Information cue labels and their display intents	113
Table 5.1 Techniques used in offline evaluation experiment	121
Table 5.2 Aggregated ranks generated using CE algorithm with spearman footrule distance	125
Table 5.3 Aggregated ranks generated using CE algorithm with kendall distance ..	125
Table 5.4 Number of participants by stage	133
Table 5.5 Participant demographic variables	134
Table 5.6 Research topics used by the participants	137
Table 5.7 Survey questions and corresponding measures for Task 1	141
Table 5.8 Independent samples t-test results of Task 1	145
Table 5.9 Spearman correlation between measures of Task 1	147
Table 5.10 Multiple linear regression results of Task 1	148
Table 5.11 Evaluation questions and corresponding measures of Task 2	149

Table 5.12 Independent samples t-test results of Task 2	154
Table 5.13 Measure combinations with moderate to high correlations of Task 2....	156
Table 5.14 Multiple linear regression results of Task 2.....	157
Table 5.15 Evaluation questions and corresponding measures of Task 3	159
Table 5.16 Independent samples t-test results of Task 3	163
Table 5.17 Measure combinations with moderate to high correlations of Task 3	164
Table 5.18 Multiple linear regression results of Task 3.....	165
Table 5.19 System-level evaluation constructs and constituent measures	166
Table 5.20 Independent samples t-test results of system-level evaluation	172
Table 5.21 Consolidated results for the agreement percentages and hypothesis	176
Table 5.22 Predictors identified in the regression models of the three tasks.....	178
Table 5.23 Crosstab of pre-study and post-study variables of Task 1	183
Table 5.24 Crosstab of pre-study and post-study variables of Task 2	184
Table 5.25 Crosstab of pre-study and post-study variables of Task 3	186
Table 5.26 Inter-coder reliability statistics	187
Table 5.27 Preferred aspects categories of the three tasks.....	188
Table 5.28 Critical Aspects Categories of the Three Tasks	193
Table 5.29 System-level categories	198
Table A.1 Criteria Counts for the Three Techniques.....	283
Table A.2 Top 20 papers from HC technique	283
Table A. 3 Top 20 papers from TPC technique	285
Table A.4 Top 20 papers from TC technique	286
Table A.5 Ranks of the three techniques	290

LIST OF FIGURES

Figure 1.1. Conceptual summary of research design	11
Figure 2.1. Ellis information seeking model (Ellis & Haugan, 1997)	18
Figure 2.2. Intervention model for researchers	26
Figure 3.1. Sub-topics of inadequate and omitted citations surveyed in Study I.....	49
Figure 3.2. Participants data by primary discipline	55
Figure 3.3. Participants data by position and experience level	56
Figure 3.4. Experience of participants in journal & conference review and writing manuscripts	56
Figure 3.5. Usage of information sources by discipline	64
Figure 3.6. Recommended articles section in ScienceDirect.....	69
Figure 3.7. Pictorial representation of the evolution of reading list in the three tasks .	72
Figure 4.1. Paper network space for TPC and TC techniques	79
Figure 4.2. Depiction of the seven features in the research paper context	83
Figure 4.3. Sequence of activities in feature measurement exercise for the shortlisted papers	83
Figure 4.4. Process flow in the AKR technique.....	91
Figure 4.5. Process flow in the IDSP technique	97
Figure 4.6. User-item matrix in IBCF algorithm	98
Figure 4.7. Process flow in the CNS technique	103
Figure 4.9. Rec4LRW system architecture	106
Figure 4.10. Reading list task screen (Task 1).....	108
Figure 4.11. Selecting seed papers before executing Task 2	109
Figure 4.12. Sample list of recommended papers in Task 2.....	110
Figure 4.13. Input options in Task 3	110
Figure 4.14. Sample list of shortlisted papers in Task 3	111
Figure 4.15. Information cue labels aside the paper titles in the task recommendations	114
Figure 4.16. Word cloud generated with the author-specified keywords of recommended papers	115
Figure 4.17. Share co-references and co-citations features in Task 2 recommendations	116
Figure 4.18. Parent cluster papers viewing feature in Task 3 recommendations.....	116

Figure 5.1. Recent papers rank aggregation comparison	126
Figure 5.2. Popular papers rank aggregation comparison.....	126
Figure 5.3. Literature survey papers rank aggregation comparison.....	126
Figure 5.4. Diverse papers rank aggregation comparison.....	127
Figure 5.5. Task evaluation questionnaire for Task 1	143
Figure 5.6. Agreement percentages for the evaluation measures of Task 1	144
Figure 5.7. Evaluation screen of the similar papers task (Task 2)	151
Figure 5.8. Agreement percentages of the evaluation measures of Task 2.....	153
Figure 5.9. Evaluation screen of the shortlisting papers task (Task 3)	160
Figure 5.10. Agreement percentages of the evaluation measures of Task 3.....	162
Figure 5.11. Agreement percentage results for Effort to use the System construct....	168
Figure 5.12. Agreement percentage results for Perceived System Effectiveness construct	169
Figure 5.13. Agreement percentage results for Perceived Usefulness construct	170
Figure 5.14. Comparison of five measures between the two LR tasks for students group	179
Figure 5.15. Comparison of five measures between the two LR tasks for staff group	180
Figure 5.16. Agreement percentages of common task measures	182
Figure 5.17. Pre-study and post-study comparison of Task 1.....	183
Figure 5.18. Pre-study and post-study comparison of Task 2.....	184
Figure 5.19. Pre-study and post-study comparison of Task 3.....	185
Figure A 1. GL vs Non-GL references percentage in ACM DL articles	278
Figure A.2. Comparison of GLAM values for the four techniques (N@10).....	281
Figure A.3. Comparison of GLAM values for the four techniques (N@20).....	281

LIST OF ABBREVIATIONS

AEIN	Active and Explicit Information Needs
AKR	Author-specified Keywords based Retrieval
C	Coverage
CB	Content Based
CC	Citation Count
CF	Collaborative Filtering
CNS	Citation Network based Shortlisting
GLP	Grey Literature Percentage
IBCF	Item Based Collaborative Filtering
IDSP	Integrated Discovery of Similar Papers
IOC	Inadequate and Omitted Citations
LR	Literature Review
MP	Manuscript Preparation
PIIN	Passive and Implicit Information Needs
RC	References Count
RE	Recency
Rec4LRW	Recommender Systems for Literature Review and Writing
RS	Recommender System
S	Specificity
TC	Topical Coverage
TPC	Topical and Peripheral Coverage
TS	Textual Similarity

UBCF Used Based Collaborative Filtering

SPRS Scientific Paper Recommender System

CHAPTER ONE - INTRODUCTION

1.1 Background

“Information-seeking must be one of our most fundamental methods for coping with our environment. The strategies we learn to use in gathering information may turn out to be far important in the long run than specific pieces of knowledge we may pick up in our formal education and then soon forget as we go about wrestling with our day-to-day problems”

-Donohew, Tipton, & Haney, (1978)

These observations from (Donohew et al., 1978) fundamentally highlight three basic information seeking characteristics of humans. The first characteristic is the intrinsic need for information in dealing with day-to-day activities, the second characteristic is the knowledge of information acquisition methods and the third characteristic is the varying relevance of the acquired information across different time periods. The information seeking behavior of humans is a continuous process where the seeking strategies evolve with time due to increased awareness of four entities - information objects, information sources, information paths and relevance criteria (Ingwersen & Järvelin, 2006). At the highest level, the aforementioned four entities are common across different domains. At the transactional level, domain-specific rules and heuristics are necessary in lieu of the operating environment (Bhavnani, 2002). In the case of scientists and academic researchers, the most sought-after information object is the scholarly paper as it contains empirical results, reviews existing results and offers opinions. Even though, scientific information can be acquired from people and grey literature (Ellis, Cox, & Hall, 1993), the validity and accuracy of information cannot

be guaranteed. Majority of the academic papers are scrutinized through the peer-review process, therefore readers can expect an adequate level of consistency. Scholarly papers hence form the core part of the academic body of knowledge (BoK).

The search for scholarly papers is performed at different stages of the research lifecycle for varying information needs. Before the advent of online information systems such as academic databases, search engines and digital libraries, physical libraries were the main source of academic literature. The digital availability of academic literature was a breakthrough event as it facilitated quicker access to large volumes of data which was generated by scanning books and journals into the digital format. Furthermore, the advent of World Wide Web (WWW) made the digital information available in a ubiquitous manner. On the other hand, information retrieval (IR) based systems brought the much-needed mechanism to query the digital information (Salton & McGill, 1983). Information needs represented by natural language keywords are the main mode of input to an IR system which subsequently retrieves the matching documents (scientific papers).

Even though the IR mechanism simplified the process of searching information, there are some associated caveats. The first caveat being the expertise factor in optimally utilizing the systems (Tabatabai & Shore, 2005) and the second being the limited input modes in the IR mechanism. The first caveat has been addressed by specialized training and assistance from librarians and experts for novice researchers (Du & Evans, 2011a; Spezi, 2016) while the second caveat has been partially addressed by extensions such as task-based IR systems (Vakkari, 2001) and semantically-enhanced IR systems (Kiryakov, Popov, Terziev, Manov, & Ognyanoff, 2004). The modal limitations of IR have also been mitigated by recommender systems (RS) since these systems can be designed to process multiple contextual dimensions as

input (Resnick & Varian, 1997). Examples of additional inputs include user profile, user activity history from access/query logs, user ratings, user's spatial and temporal data, to name a few. These dimensions help in offering personalized recommendations to different users. Apart from the flexibility in accepting different types of input, a wide variety of algorithms can be used for generating recommendations using RS. Both IR and RS are part of the information filtering (IF) paradigm (Hanani, Shapira, & Shoval, 2001) where the objective is to filter the information as per user's specific information need(s). In fact, RS research subsumes IR research since content-based RS techniques involve the use of textual similarity algorithms that simulate the IR mechanism.

In the context of scientific information seeking, IR mechanism has been prevalent in academic information systems while RS techniques have been used on a complimentary basis for certain sub-tasks such as recommending similar papers for a given paper. After the inception of citation indices (Giles, Bollacker, & Lawrence, 1998), academic IR systems have incorporated bibliometric enhancements so that the citation networks could be mined for providing more relevant research papers (Ding, Chowdhury, Foo, & Qian, 2000; Wolfram, 2003). Therefore, bibliometrically-enhanced IR (BIR) is a hybrid combination of textual and non-textual (citation networks) retrieval techniques. Citation networks are an important source of information for RS as well. There have been calls to incorporate scientometrics (Mutschke & Mayr, 2015) and natural language processing (Wolfram, 2016) models so that quality of the retrieved/recommended results could be improved. The topic of recommending papers for researcher's information needs from both IR and RS perspectives has organically evolved into an interdisciplinary and multidisciplinary research topic.

1.2 Motivation

RS techniques have been conceptualized to recommend a wide variety of scholarly information objects such as publication venues (Yang & Davison, 2012) and collaborators (Gunawardena & Weber, 2009) in addition to the standard scientific papers (Mcnee, 2006). These information objects are required at different stages in the research lifecycle. In particular, scientific paper recommender systems (SPRS) research has been an active area. SPRS techniques from prior studies (Bae, Hwang, Kim, & Faloutsos, 2014; Ekstrand et al., 2010; Liang, Li, & Qian, 2011) provide relevant papers for the corresponding researcher search tasks. These techniques utilize data from sources such as the citation network, paper metadata, full-text and system usage log files for generating recommendations. The selection of algorithm for these techniques is based on the task requirements. For instance, graph ranking algorithms are used in (Bae et al., 2014; Ekstrand et al., 2010) for recommending popular papers while custom designed citation linking technique is used in (Liang et al., 2011) for recommending similar papers for a given seed paper. However, the diversity of the techniques makes it a difficult proposition for in-house implementation in an environment where multiple search tasks are supposed to be supported. Each technique needs to be separately implemented along with the different set of fields and features depending on the input requirements of the corresponding recommendation technique. The diversity in the datasets and the nature of the incumbent algorithm (machine-learning vs. classical information filtering) are other factors that affect the implementation of piecemeal approaches from unconnected studies. This scenario creates the need for an intermediate framework or structure for connecting the researchers' search tasks to the RS techniques.

The second drawback of existing approaches is the lack of interconnectivity between the recommendation techniques for tasks which are executed in a sequence. Examples of serial tasks typically include finding papers for literature review and finding topically similar papers, to name a few. The researchers' selection of research papers evolves through these tasks in a natural setting. Therefore, RS can be more effective with task interconnectivity as a prerequisite for multiple tasks. This issue has already been observed and highlighted in the context of IR systems (Huurdeeman & Kamps, 2014; Huurdeeman, Wilson, & Kamps, 2016).

An unexplored area in this research area is the use of article-type as one of the contextual dimensions for generating citation recommendations during the manuscript preparatory stage of the research lifecycle. Article-type ranges from journal survey/review papers, journal case studies to conference long papers and short papers. Article-type has been highlighted as part of the user's 'Purpose' in the multi-layer contextual model related to SPRS (Dehghani, Afshar, Jamali, & Nematbakhsh, 2011). Since the article-type preference is part of the user's goal, the retrieval/recommendation process should consider this dimension for identifying papers that are closer to user's requirements. It is expected that the user satisfaction will increase when resources are recommended based on higher number of contextual dimensions. Thus, RS techniques can be used to provide recommendations based on article-type preference of the user during the manuscript writing/preparation stage.

1.3 Research Objectives

The overarching goal of this research was to build an assistive system using IR/RS techniques for helping researchers in finding papers for selected set of key literature review (LR) and manuscript preparatory (MP) tasks. For achieving this goal, two

research objectives were proposed. The first objective (RO1) was to identify an appropriate method to map the identified LR and MP tasks to relevant IR/RS algorithms. This objective was meant for the activities pertinent to the design and development of the assistive system. As a part of RO1, the current research aimed to address the following research question (RQ1):

RQ1: What are the key search tasks of researchers in the literature review and publication lifecycle?

Although the solution to RQ1 might appear minimally significant, it is important to establish the varying criticality of the tasks so that future research studies could be appropriately prioritized. RQ1 was addressed through Study I. The main topic for Study I was “Inadequate and Omitted Citations (IOC) in Research Manuscripts”. Data was collected at the problem level i.e. at the IOC level so that the appropriate solutions could be subsequently put forth. Study I was a survey based study since the research question was a suitable candidate for addressing through a quantitative approach. The study was also meant to collect data pertaining to the design of the assistive system along with user awareness of existing systems and their recommendation features. The following questions were explored as part of Study I.

- 1) What are the critical instances of IOC in research manuscripts submitted for review in journal and conferences?
- 2) Do the critical instances and reasons of IOC in research manuscripts relate with the scenarios/tasks where researchers need external assistance in finding papers?
- 3) Identify the prominent information sources that include academic databases, search engines and digital libraries, through frequency of usage.

- 4) What is the researchers' awareness level of available recommendation services for research papers?

The next research question RQ2 under RO1, dealt with the design of the recommendation techniques for the identified tasks from the earlier research question RQ1.

RQ2: How to relate the identified tasks of researchers to IR/RS algorithms?

The deliverable for RO1 was a prototype assistive system. The recommendation techniques for the identified tasks in the system were based on theoretical requirements of each task. The theoretical requirements were identified from prior literature. The identified requirements for the tasks were meant to be highlighted in the system so that users are able to perceive the difference in the recommendations offered by the system.

The second research objective (RO2) was to evaluate whether the performance of the proposed recommendation techniques for the tasks and the overall system were at the expected level. The evaluation was carried using two serial approaches. First, offline evaluation experiment was conducted for the applicable recommendation techniques with the objective to check whether the proposed techniques outperformed the baseline approaches. Second, a user evaluation study was conducted to elicit researchers' opinion of the task recommendations and the overall prototype system. The corresponding research questions RQ3 and RQ4 for RO2 were as follows:-

RQ3: Do the proposed recommendation techniques of the relevant tasks outperform the existing baseline approaches in system-based evaluation?

RQ4: Do the proposed recommendation techniques and the overall system meet the expected standards in user-based evaluation?

RO2 was addressed through Study II. For RQ3, a pre-study in the form of an offline evaluation experiment was performed. For RQ4, an extensive user evaluation study was conducted as part of Study II with researchers who had manuscript authoring experience. Participants were differentiated by experience level and current position for data analysis. The specific evaluation goals for Study II were as follows.

- 1) Ascertain the agreement percentages of the evaluation measures for the three tasks and the overall system and identify whether the values are above a preset threshold criteria of 75%.
- 2) Test the hypothesis that students benefit more from the recommendation tasks/system in comparison to staff.
- 3) Measure the correlation between the measures and build a regression model with 'agreeability on a good list' as the dependent variable.
- 4) Track the change in user perceptions between the three tasks.
- 5) Compare the pre-study and post-study variables for understanding whether the target participants are benefitted from the tasks.
- 6) Identify the top most preferred and critical aspects of the task recommendations and the system using the subjective feedback of the participants.

1.4 Overview of Research Design

As shown in the conceptual summary of the research design in Figure 1.1, this research was split into three main sequential phases. The logical beginning of Phase 1 was the review of literature. As of 2013, there were no survey papers that synthesized

the application of RS in research lifecycle related tasks. Therefore, the review of prior literature led to a realization that RS techniques were used across different stages in the research lifecycle. However, there were no standardized frameworks, evaluation approaches and datasets used in these studies. Secondly, most of the proposed techniques were not implemented in the contemporary databases, search engines and digital libraries. Hence, the research community appeared to be largely unaware of specialized RS implementations that could potentially assist them in their literature search sessions. Data regarding the awareness level of researchers on the various available systems was thus set to be collected during Study I. RO1 was partly addressed through the Study I survey in Phase 1. This survey helped in finalizing the three LR and MP tasks for the prototype, with few inputs for the recommendation techniques. Secondly, the survey findings helped in the design of the user-interface (UI) of the system.

Phase 2 was the non-empirical phase of this research. This phase encompassed the technical development activities related to the precomputation of base features, design of recommendation techniques and front-end development of the prototype. The tasks for the assistive prototype system identified from Study I were the key inputs for this phase. RQ2 and overall RO1 were addressed in this phase. There were three major activities in this phase. The first activity was the identification of the base features that relate the tasks to the RS algorithms. The second activity was the design of the recommendation techniques while the third activity was the design and development of the assistive system. The overall process in this phase was guided by a novel interventional framework comprising of three components - task redesign (for aligning the task requirements to the recommendation techniques), task interconnectivity (for addressing the management of papers within the system) and

informational display features in the user-interface (for simplifying users' interactions with the system).

The developed system from Phase 2 was subsequently evaluated in Phase 3. RO2 was addressed in Phase 3. The penultimate step in this phase dealt with the analysis of the collected evaluation data using both quantitative and qualitative techniques. Study II was designed for this purpose. The three phases of this research resemble the design science research process model (Takeda, Veerkamp, & Yoshikawa, 1990) which comprises of five steps – (i) awareness of problem, (ii) suggestion (research ideas), (iii) development, (iv) evaluation and (v) conclusion (dissemination). The approval document from the university's institutional review board (IRB) for the studies conducted as part of the current research is available in Appendix A.

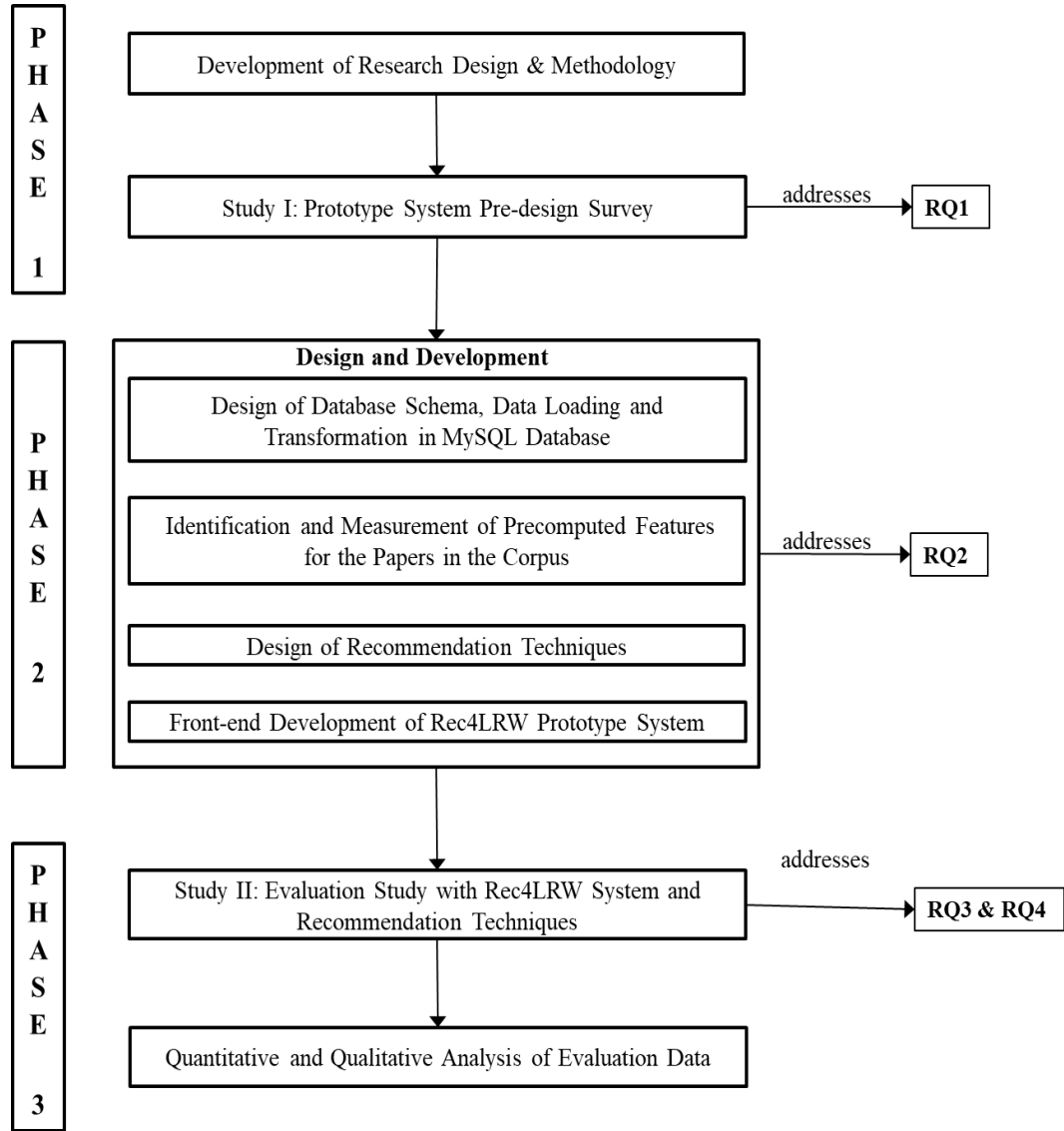


Figure 1.1. Conceptual summary of research design

1.5 Assumptions

Through the literature review conducted for this research, two types of helpful interventions for researchers were identified. These interventions aid researchers in circumnavigating the experience gap in terms of literature review and manuscript preparatory/writing skills. They are *people-oriented* and *technology-oriented* interventions. The former involves manual assistance from experts and librarians (Du & Evans, 2011b; Spezi, 2016) while the latter involves deployment of specialized

information systems for helping researchers (Atanassova & Bertin, 2014; Chou & Yang, 2011; Gamberini et al., 2015). The current research assumes that researchers would use the system developed as a part of this research, along with the other interventions proposed from prior studies. The current research doesn't seek to replace *people-oriented* interventions with a new system. The importance of *people-oriented* interventions is well-known in the academic community as such interventions are part of the traditional method to help novice researchers. Secondly, the dependence on information sources such academic databases and digital libraries cannot be completely replaced by the assistive system due to two reasons. First, these systems are the primary sources of information where the publications data are indexed. Second, these systems are constantly improving the UI features so that users are able to perform better searching and browsing. The main drawback of these systems is that they are more suited for ad-hoc search tasks. Even though, the assistive system developed as a part of this research is dependent on data from these primary sources, the expectation is that the corpus will be updated with data from federated sources on a periodic basis. The ultimate goal of this research is to provide confidence to database and digital library administrators in considering a task-based interface so that the proposed recommendation techniques could be incorporated.

1.6 Dissertation Organizational Overview

This dissertation is organized in the following manner. Chapter 2 presents a literature review of the areas - scientific information seeking studies and the use of RS across the scientific research lifecycle. In Chapter 3, Study I results are presented. In Chapter 4, the design and development details of the Rec4LRW system are outlined including the base features and the recommendation techniques along with other implementation details. In Chapter 5, the results of the offline experiment and user evaluation study are

presented as part of Study II. In Chapter 6, the concluding remarks, contributions, implications, limitations and future works are discussed.

CHAPTER TWO - LITERATURE REVIEW

2.1 Introduction

The literature review chapter provides an introduction to the relevant studies in research topics where the focus has been directed towards helping researchers in finding papers for their literature review and manuscript writing/preparatory purposes. This chapter is organized in two parts. In Part One, scientific information seeking models and studies are discussed. Part Two is dedicated to recommender systems (RS). The section covers basic concepts and review of studies that are about the application of RS in scientific research lifecycle tasks.

2.2 Part One - Scientific Information Seeking

The scientific research lifecycle encompasses the activities performed by researchers across different disciplines. There have been multiple versions of this lifecycle put forth in previous studies. Björk & Hedlund (2003) used the term “Scientific Publication Lifecycle (SPLC)” for referring to the lifecycle of a single publication (journal article) with activities ranging from the actual research to paper writing and then onto the readership and practical implementations. The term “Scholarly Communication Lifecycle” has been loosely used in certain studies (Harley, Acord, Earl-Novell, & Lawrence, 2010; Wright, Sumner, Moore, & Koch, 2007) to highlight the different avenues of publication such as traditional journals, open-access journals and pre-print services.

The research lifecycle schematic was put forth in (Nicholas & Rowlands, 2011) for collecting data on social media use in the research workflow. The eight steps in this schematic are an adequate representation of the lifecycle. The eight steps are (i)

Identify research opportunities, (ii) Find collaborators, (iii) Secure support, (iv) Review the literature, (v) Collect research data, (vi) Analyze research data, (vii) Disseminate findings and (viii). Throughout this lifecycle, researchers seek information for various information needs. Johnson (1997) defines information seeking as “purposive acquisition of information from selected information carriers”. Apart from research papers, researchers would be interested in seeking information pertaining to other scholarly information objects such as patents, publication venues, funding agencies, collaborators and employment opportunities to name a few.

2.2.1 IS Models Overview

In the research area of information seeking (IS), models have been put forth as a precursor to subsequent theory development (Bates, 2005) and to also serve as a framework for thinking about a problem area (Wilson, 1999). These models can be purely conceptual models that need to be tested/validated through empirical data (Bates, 2005) or they could be generated based on empirical findings (e.g. Ellis model (Ellis et al., 1993)). Savolainen (2015) notes that IS models tended to cover cognitive dimensions while lesser focus has been given to affective dimensions such as emotions and feelings. The selection of dimensions is based on the domain and the main intent behind the conceptualization of the model. As for the intent, Johnson (1997) has identified three purposes behind development of IS models. They are (i) IS models should provide a theoretical basis for predicting IS behaviors, (ii) IS models should provide strategies for enhancing IS, and (iii) IS models should provide answers for why users pursue particular IS patterns.

IS models are either generalistic models or domain-specific models.

Generalistic models are domain-agnostic models that are abstract and applicable to

most information seeking scenarios. Few examples of generalistic models include (i) Krikelas model (Krikelas, 1983) which was designed as a daily life model focusing on “information gathering” and “information giving” of people, (ii) Johnson model (Johnson, 1997) which differentiates itself by including antecedent factors for the process of information seeking and (iii) Savolainen model (Savolainen, 1995) which is also called the Everyday Life Information Seeking (ELIS) model, specifically meant for casual (non-work) information seeking tasks. Domain-specific models have been conceptualized for individual domains with the aim of capturing particulars specific to the domains. Examples of such models include Ellis model (Ellis et al., 1993) which is for scientists and PRISM model (Kahlor, 2010) which is related to health information seeking. Apart from generalistic and domain-specific models discussed in the previous paragraph, there are more binary classifications put forth by Ingwersen & Järvelin (2006). They are (i) broad vs. narrow, (ii) process vs. static, (iii) abstract vs. concrete, and (iv) summary vs. analytical models.

IS model research has progressed in a direction where integration has been a central theme whenever the context has moved beyond a particular frame of reference. For instance, there are integrated models of information seeking & searching (Bates, 2002), integrated information seeking & retrieval framework (Ingwersen & Järvelin, 2006) and integrated information behavior model (Lakshminarayanan, 2010). Thus, IS models offer the necessary substratum for subsequent applied research which are meant to address particular issues.

2.2.2 Scientific IS Models

In the context of the current research, scientific information seeking is the relevant area of information seeking. One of the aims of scientific information seeking research

is to identify factors and model the sequence of steps used by researchers in seeking information related to their academic needs. The work of Bichteler & Ward (1989) was one of the earliest works with specific focus on the problems faced by geoscientists. The study results indicate the importance of the role of librarians since the decision making process involved in selecting resources requires expertise. The role of librarians cannot be understated as they frequently play the role of a *cognitive authority* for researchers and general public (Wilson, 1983). The study of Palmer (1991) looked at the influence of variables such as personality, discipline and organizational structure on the information seeking behavior of researchers involved in agriculture related research. The study results indicate the importance of having multiple and diverse information sources for satisfying information needs.

Ellis's multiple studies (Ellis, 1993; Ellis et al., 1993; Ellis & Haugan, 1997) in scientific information seeking are one of the most important contributions in this area. The eight activities in the Ellis model are *starting*, *chaining*, *browsing*, *differentiating*, *monitoring*, *extracting*, *verifying* and *ending* (as shown in Figure 2.1). *Starting* refers to the activity where the researcher starts an initial search for the required information and it roughly corresponds to the zooming literature search technique (Levy & Ellis, 2006). *Chaining* activity is split into two sub-activities forward chaining and backward chaining where the researchers follow the backward references trail and forward citations network to find relevant papers respectively. The combined process of forward and backward chaining has been referred to as "intellectual structure" (White & Griffith, 1981). *Chaining* corresponds to concertina literature search technique (Levy & Ellis, 2006) since researchers zoom in on a topic followed by expanding the search with more papers. *Browsing* activity is the stage where the

researcher identifies potential area of interest and conducts semi-directed search. This activity corresponds to the lens searching technique (Levy & Ellis, 2006).

Differentiating refers to the activity in which the information sources are tweaked to observe differences in the obtained information. *Monitoring* is a periodic activity meant to track new updates in the field or topic of interest. *Extracting* is the activity where selective information is extracted from the information sources as per the need in hand. The final two activities *Verifying* and *Ending* were added at a later stage (Ellis & Haugan, 1997) while the first model (Ellis, 1993) that was conceptualized based on social science researchers ended with the extraction activity itself. *Verifying* is an activity that is particularly applicable for researchers dealing with numerical data which presents likelihood for errors. *Ending* represents the final stage of information seeking where researchers put a logical end to the task in hand. It is to be noted that activities such as *Chaining*, *Browsing* and *Differentiating* are activities analogous to the literature search strategies particularly concertina, lens and zooming (Levy & Ellis, 2006; Ridley, 2012). The activities in this model closely mirror the steps followed in performing literature review particularly in the stages - searching and reading (Ridley, 2012).

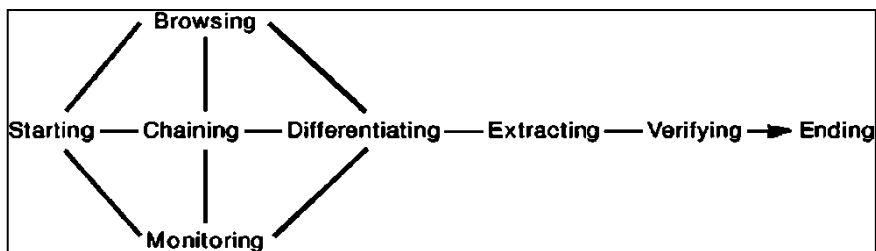


Figure 2.1. Ellis information seeking model (Ellis & Haugan, 1997)

Chowdhury *et al.* (2014) put forth a recent IS model based on data collected from researchers. In this model, uncertainty is taken as the frame of reference in the

context of information seeking & retrieval (IS&R). The main finding of this study is that it challenges the notion that the level of uncertainty decreases as the user progresses through the information seeking process. Instead, the study provides evidence for the persistence of uncertainty in different stages of information seeking. Since the data for this model has been collected from the information seeking perspective of research students and staff, it can also be perceived as a scientific IS model. The model consolidates relevant IS activities and IS problems of researchers (as listed in Table 2.1). The model differentiates uncertainty into (i) negative type which leads to anxiety and lack of confidence and (ii) positive type which leads to new ideas and serendipitous information, thereby eventually leading to new ideas for research towards the end of the IS process.

Table 2.1

IS activities and problems identified in Chowdhury et al. (2014)

Information Seeking Activities	Information Seeking Problems
Choosing Information channels/sources	Information overload
Formulating a search expression	Unaware of source/channel
Deciding how many and which items to view	Search output is not exhaustive
Deciding when to stop a search and begin to use items	Unfamiliar with the source search results are not up-to date
Judging the quality of sources	Information is too scattered
Making sure to remain up-to date in field	
Ensuring that all the information	

required for a task has been obtained

2.2.3 Scientific IS Behavior Studies

In this sub-section, behavior studies conducted in the area of scientific information seeking are discussed with a focus on specific sub-topics.

2.2.3.1 Issues Faced by Researchers during Scientific IS

Barrett (2005) studied the information seeking habits of graduate students in a humanities school. Findings of this study showed that the individual sessions of graduates tend to be idiosyncratic with projects started in a haphazard manner. Results indicated student's reliance on supervisors' guidance in finding relevant documents, thereby underlining the complimentary role of intelligent systems in addition to online library services such as Online Public Access Catalog (OPAC), in supporting students' information needs. Barrett's study is in-line with an earlier study conducted by Fidzani (1998) which highlights the graduate students' inability in using the library resources in a required manner to perform research. The issue of information overload is also stated as one of the reason for graduate students' inability in performing a proper synthesis of literature (Switzer & Perdue, 2011). This is a situation where the volume of available information is too high for students in properly synthesizing the state-of-the-art in the particular research area. Information overload could lead to the anxiety among students which further exacerbates the problem (Needham, 2007). It is to be noted the information overload can lead to poor decision quality due to people's limited cognitive processing capacity in lieu of excess information (Speier, Valacich, & Vessey, 1999). The assistive role required from library resources and librarians in helping researchers is a common observation in many studies (Catalano, 2013; George et al., 2006; Head, 2007; Wu & Chen, 2014). However, there have been cases where

students have been dissatisfied with library instructions, creating a need for further flexibility in personalizing the instructions (Blummer, Watulak, & Kenton, 2012).

The study conducted by Du & Evans (2011) identified characteristics, strategies and sources related to research-related search tasks of graduate students. The study indicates the difficulties faced by researchers in finding relevant information using multiple explorative search sessions in multiple sources. Spezi (2016) pointed out the lack of depth in the skills of doctoral students in ascertaining relevance of search results and overall use of academic information sources such as search systems and digital libraries. Low system skills (Bullock, 2013) and lack of ICT confidence (Markauskaite, 2007) are other associated deficiencies with students. The nuances in handling the features provided by academic search systems are highlighted with marked differences between novices and experts. This issue is also echoed in other studies (Brand-Gruwel, Wopereis, & Vermetten, 2005; Karlsson et al., 2012; Tabatabai & Shore, 2005; Yoo & Mosa, 2015) where the experts' ability in carefully formulating a problem before conducting search is highlighted as a key difference. Academic experience and searching competency are two factors that affect the confidence of researchers, thereby differentiating expert researchers from novices (Niu & Hemminger, 2012).

Studies related to exploratory search and search tasks have identified the characteristics of the search tasks undertaken by researchers and attributed the complexity of these tasks to characteristics such as uncertainty and prior knowledge of researchers (Liu & Kim, 2013; Wildemuth & Freund, 2012). However, the search tasks are not always caused by a set of correlated reasons, instead the reasons are independent, as per a study (Liu, 2015) conducted with undergraduate students. In addition, there are also limitations with the user-interfaces (UI) that can impede simple

search tasks (Diriye, Blandford, Tombros, & Vakkari, 2013). Such limitations are exacerbated by the finding that seeking and search behavior varies significantly by the type of the task (Byström & Järvelin, 1995; Liu et al., 2010). Therefore, there are multiple reasons that affect and impede researchers during the information seeking sessions.

2.2.3.2 Remedial Approaches

Most of the prior studies suggest general remedial measures such as educating new users on information seeking best practices, involving librarians for assisting users and involving experts to assist novices so that their information seeking, searching skills and competencies improve. It is important to classify the remedial approaches or interventions. At a high level, these interventions can be categorized to two types - *people-oriented* and *technology-oriented* interventions. The importance of the librarians and expert researchers has been highlighted in helping rookie researchers (Du & Evans, 2011b; Spezi, 2016). Topic-specific protocol put together by experts is another proposed approach in (Curró & Pretto, 2016) where search terms, journals list and organizational structure were provided to graduate students. The importance of metacognitive search behavior has been stressed in education research dealing with undergraduates, since it has been related to better task performance (She et al., 2012). Metacognitive knowledge is basically knowledge of self, the current tasks and the search strategies (Wolf, Brush, & Saye, 2003). However, metacognition based remedial approaches such as self-learning, planning and monitoring cannot be regarded as final remedies since there is additional mental load.

On the other end of the interventions spectrum, *technology-oriented* interventions from prior studies include search systems with faceted user interfaces for

better display of search results (Atanassova & Bertin, 2014), stage-specific search user-interface features (Huurdeeman et al., 2016), symbiotic systems (Gamberini et al., 2015), meta-search systems and engines (Hoeber & Khazaei, 2015; Sturm, Schneider, & Sunyaev, 2015), personalized search (Salehi, Du, & Ashman, 2015), bibliometric tools for visualizing citation networks (Chou & Yang, 2011) and scientific paper recommender systems in both embedded mode (Joeran Beel, Langer, Genzmehr, & Nürnberger, 2013; Naak, Hage, & Aïmeur, 2008) and stand-alone mode (Huang, Wu, Mitra, & Giles, 2014; Küçüktunç, Saule, Kaya, & Çatalyürek, 2013).

2.2.4 Summary of Part One

In the area of scientific information seeking and academic search tasks, research has progressed in a direction aimed at understanding the activities, related heuristics and issues faced by researchers. Models such as Ellis model (Ellis & Haugan, 1997) and (Chowdhury et al., 2014) have been put forth to give an overview of the activities and scenarios involved in scientific information seeking. Information behavior studies have been conducted to understand the heuristics and difficulties faced by researchers during information seeking. These studies point out the differences between researchers by associating the contextual variables (e.g., prior knowledge and task complexity, uncertainty). These studies unanimously highlight that novice researchers and researchers venturing into new areas, have difficulty in finding information sources and identifying the appropriate resources for their information needs.

Remedial solutions have been provided through *people-oriented* and *technology-oriented* interventions. From the people front, the role of supervisors and librarians in assisting novice researchers has been most repeatedly made along with calls for specialized training. From the technology front, intuitive user interfaces, specialized search systems, bibliometric tools have been recommended. It can be claimed that

current academic systems need to provide features to help researchers with different information seeking tasks during the entire lifecycle of scientific research. It is to be noted that researchers make use of both these interventions in practical scenarios. These interventions are mutually exclusive only in terms of definition since a novice researcher avail the service of librarians and senior researchers even while learning to use academic information systems.

In Figure 2.2, the intervention model for researchers is depicted. This model is just for illustration purposes, meant to visually highlight the fundamental differences between two interventions. It is to be noted that this model doesn't attempt to disincentivize the usage of *people-oriented* interventions. This model is applicable in the context of researchers searching for papers during literature review and manuscript preparation. This model mostly applies to the case of novice researchers who have both low knowledge and low experience in terms of information seeking. There are two levels in the model namely the current level and expected level. The current and expected levels are different for novices while for most experts, the levels are indistinguishable due to their experience. The objective for novices is to reach the expected level from their current level i.e., find research papers for their requirements. The ideal scenario is *Path A* since it represents a straight line connecting two points, which is the shortest distance between two points. Since novices don't have the required knowledge and experience, they have to rely on certain interventions to reach the expected level. The traditional interventions are *people-oriented* where researchers seek the help of research supervisors, subject-matter experts, librarians and colleagues. Even though, these interventions are useful and ideal, they are affected by temporal and resource constraints. These individuals are not available at all times and secondly,

the novice researcher might be in an isolated location. Hence, *Path B* (from Figure 2.2) that represents the *people-oriented* interventions forms a circular path.

The second type of interventions is the *technology-oriented* interventions. These interventions are facilitated by stand-alone and embedded recommender systems, certain reference management systems such as Docear (Joeran Beel, Langer, Genzmehr, & Nürnberger, 2013), Mendeley (Vargas, Hristakeva, & Jack, 2016) and other assistive systems. Such interventions are mostly web-based information systems accessible throughout the year to novices or researchers of any experience level. These systems provide a shortcut to novices by suggesting/recommending the “right” types of papers which the novices should read. These interventions are represented by *Path C* in Figure 2.2 where there is an intermediate level (point) that connects the current and expected levels. *Path C* is much easier path for novices when compared to the traditional *Path B*. There are caveats associated with *Path C*. Factors such as inconvenient system design, unfamiliarity with the system and outdated/inflexible paper recommendation algorithms could very well hamper the experience of researchers with these systems. However, the long-term intention of both paths *B* and *C* is to elevate the experience level of novices so that *Path A* becomes the norm for them. Through this model, an attempt has been made to differentiate the two types of interventions.

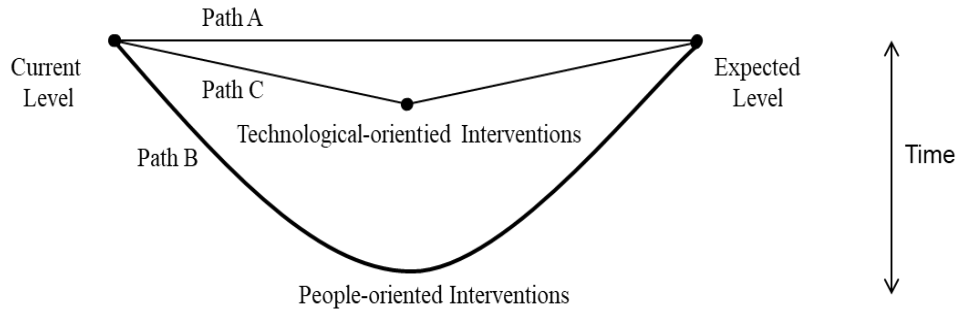


Figure 2.2. Intervention model for researchers

2.3 Part Two - Scientific Paper Recommendation Systems

2.3.1 Introduction

There have been multiple definitions put forth on recommender systems (RS) in the context of personalization and information overload (Eppler & Mengis, 2004). Burke (2002) defines RS as “any system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible objects”. The view of Mcnee (2006) is in a similar vein with RS seen as entities to help users in finding items in a complex information space. Xiao & Benbasat (2007) define RS as “software agents that elicit the interests or preferences of individual user for products, either explicitly or implicitly and make recommendations accordingly”. The keyword “products” in the definition can be replaced with resources or items for applicability in all domains. Burke’s definition is apt for this study as it is a comprehensive definition.

An initial variant of RS was first introduced in 1992 with Tapestry, a system meant to handle large collection of email messages (Goldberg et al., 1992). At that stage, the technique was academically referred to as collaborative filtering (CF) which continues to remain as one of the most used RS algorithm till date. As research

continued in CF, there was an important study that looked at both similarities and differences between information retrieval (IR) and information filtering (IF) systems by Belkin & Croft (1992). Information filtering systems are contemporarily represented by content based recommender systems (CB) (Lops, Gemmis, & Semeraro, 2011) which provide some flexibility over the traditional IR systems by making recommendations based on multiple criteria.

2.3.2 Classification of RS

The broad nature of RS research is best showcased by the wide variety of techniques available for producing recommendations based on the application context.

Collaborative filtering (CF), content based (CB) and hybrid RS are the most accepted and adopted techniques (Jannach, Zanker, Felfernig, & Friedrich, 2010; Mcnee, 2006a; Parra & Sahebi, 2013) while there are other types such as knowledge based (KB) RS (Jannach et al., 2010) and Social RS (Lü, Medo, Yeung, & Zhang, 2012).

Apart from the aforementioned standard RS techniques, there are other techniques such as multi-criteria RS and contextual RS that are relevant to many environments. Traditional recommendation techniques are based on a single criterion by default. In situations, where the system is able to collect multi-dimensional or multifaceted feedback from the user, there is availability of multiple criteria ratings for resources. The type of RS that operate on multiple criteria are called multi-criteria RS (Manouselis & Costopoulou, 2007). Contextual RS draw their line of tradition from contextual search (IR) systems (Pasi, 2011; Pasi, Informatica, & Bordogna, 2007) where research has been focused on identifying contextual dimensions surrounding a user's search session so that most relevant resources are provided back to the user. Of late, deep learning (LeCun, Bengio, & Hinton, 2015) approaches have been

implemented in SPRS for providing textual recommendations (Bansal, Belanger, & McCallum, 2016). The scope for combining recurrent neural networks (Werbos, 1990) with CF technique has been explored in these studies.

2.3.3 Usage of RS in Scientific Research Lifecycle

RS techniques have been proposed to recommend information objects for a wide variety of tasks in the research lifecycle. In this domain, RS is known in different names such as research paper RS (Beel, Langer, Genzmehr, Gipp, et al., 2013), scientific paper RS (Sesagiri Raamkumar, Foo, & Pang, 2015), citation RS (Huang et al., 2014) and literature RS (W. Yang & Lin, 2013). With about 250 papers published in this area with the earliest work from Basu et al. (2001) in 2001, there have been different categorization schemes used in previous studies. Beel & Langer (2014) categorized the papers using seven concepts which are largely about the algorithm types. Champiri, Shahamiri, & Salim (2015) categorized RS papers based on nine types of contextual information. These two classification schemes are listed in Table 2.2.

Table 2.2

Categorization schemes from previous studies of SPRS

Classification of Beel et al., 2014	Classification of Champiri et al., 2015
1. Stereotypes	1. Profile (Long term/short term profile)
2. Content Based Filtering	2. Types
3. Collaborative Filtering	3. Purpose/target
4. Item-Centric / Co-Occurrence Based	4. Activity/task
5. Graph-based	5. Pre-knowledge/skill

Classification of Beel et al., 2014	Classification of Champiri et al., 2015
6. Global Relevance	6. Social networking & homepage
7. Hybrid	information
	7. Logs
	8. Information behavior
	9. Information need level

In the current review, the categorization scheme is meant for covering the entire research lifecycle, thereby extending beyond studies related to SPRS. Terminologies from information seeking behavior (ISB) have been used to categorize the studies. The categorization scheme has a multi-level structure starting with two high level categories. They are (i) active and explicit information needs and (ii) passive and implicit information needs. These categories are mutually exclusive in terms of classification. The categorization scheme used in the current review is listed in Table 2.3.

Table 2.3

Categorization scheme for application for RS in research lifecycle tasks

Active and Explicit Information Needs (AEIN)	Passive and Implicit Information Needs (PIIN)
1. Identify research opportunities	1. User Footprint
2. Find collaborators	2. Researcher's Publication History
3. Review the literature	3. Social Network of Authors
3.1. Task of building an initial reading list at the start of LR	4. Social Tags
	5. Periodic Recommendations from

Active and Explicit Information Needs (AEIN)	Passive and Implicit Information Needs (PIIN)
3.2. Task of finding similar papers based on a single paper	Reference Management Systems
3.3. Task of finding similar papers based on multiple papers	
3.4. Task of searching papers based on input text	
4. Disseminate Findings	
4.1. Publication Venues	
4.2 Citation Contexts	

2.3.3.1 Active and Explicit Information Needs (AEIN)

The AEIN category represents the set of studies where the information needs of researchers are both active and explicit. Active in the sense that the researchers are actively looking for research papers and explicit because they express their information needs directly to the information systems through recognizable input modes such as search keywords and seed paper(s). Under this AEIN category, the sub-categories are allocated based on the stages in the research lifecycle put forth by Nicholas & Rowlands (2011). It is to be noted that prior studies have not applied RS for all the eight stages of the lifecycle. The stages which have not been addressed thus far in RS studies are stage 3 - secure support, stage 5- collect research data and stage 8 - manage the research process.

Identify Research Opportunities

This is the initial stage in the lifecycle where researchers are on the lookout for interesting research areas which provide scope for novel contributions. The corresponding requirement is to find research papers. There are not many RS studies that have concentrated on this stage. Identifying research papers for this requirement can be considered to be more of a big data problem since it needs technical foresight (Hahn, 2016). This requirement was partially addressed by one of the information seeking tasks - “Explore research interest” in (Mcnee, 2006). As per the study results, user-based CF (UBCF) was an RS algorithm best suited for the purpose. The model proposed by (Suntharasaj & Kocaoglu, 2013) can be used in future RS studies for identifying research opportunities.

Find Collaborators

During this stage, researchers intend to find potential collaborators from other disciplines to work on interdisciplinary and multidisciplinary projects. RS techniques have been proposed for this stage in (Gunawardena, 2013; Gunawardena & Weber, 2009) where a case-based reasoning (CBR) methodology was used. Specifically, a combination of clustering (Manning, Raghavan, & Schütze, 2008) and association rules techniques (Agrawal et al., 1993) was used in the formulation of recommendations based on positive and negative indicators of past collaborations.

Review the Literature

The LR stage has been well addressed by RS studies. Majority of the RS studies are potentially SPRS studies applicable for implementation as LR recommendation tasks. This stage has another level of subcategories. These subcategories are represented by

LR search tasks which are practical information seeking tasks. The studies related to these tasks are discussed as follows.

Task of Building an Initial Reading List at the Start of LR

A reading list is defined as “a list of sources (recommended by a teacher or university lecturer) which provide additional or background information on a subject being studied” (Collins, 2016). In the context of LR, there have been no formal definitions set for this collection of papers. Papers considered as seminal, classical or important in a particular research area, have constituted the reading list in previous studies. The advent of collaborative filtering (CF) algorithms (Goldberg et al., 1992) have benefitted this area of research. Basic versions of CF algorithms have been found to produce better results than content-based (CB) filtering methods for multiple LR search tasks (Mcnee, 2006). Continuing the trend, CF techniques weighted with scores from graph ranking algorithms PageRank, Hyperlink-Induced Topic Search (HITS) and Stochastic Approach for Link-Structure Analysis (SALSA) provided even better results (Ekstrand et al., 2010) than traditional CF algorithms.

Two unique characteristics behind the citing behavior of classical (seminal) papers were used to conceptualize an approach for identifying seminal papers (Wang, Zhai, Hu, & Chen, 2010). These two characteristics are (i) the nature of seminal papers being consistently downloaded for reading and (ii) the habit of researchers in citing references from a classical paper. A paper ranking score was calculated by operationalizing the two characteristics. Novelty of a research paper towards the particular research area was the focus of a study where the citation network was leveraged to identify critical papers (Chen et al., 2011). A methodology known as citation authority diffusion (CAD) was proposed to identify important papers from

Google Scholar for particular research topics. This approach is reliant on set of input research papers for identifying the relevant papers. Bae et al. (2014) proposed an algorithm based on random walk with restart (RWR) algorithm to measure the seminality score of papers. The method incorporates inter-paper similarity scores in addition. However, this approach is aimed at building a genealogy of seminal research papers so that researchers are able to track the progress of particular trends. Another recent study employed a hybrid approach by combining topic models with PageRank. The proposed ThemedPageRank algorithm has been evaluated to provide better results than Google Scholar (Jardine, 2014). This approach also takes the age of research papers into account.

There are certain shortcomings in these studies. Firstly, it is important to note that the utility of these methods have not been tested in the interdisciplinary research sphere. Secondly, there are certain limitations such as lack of semantic relations between citations (Hjørland, 2013) in all the studies surveyed as part of the related work. Thirdly, these approaches assume that seminality of a research paper is the main characteristic for inclusion in a reading list. When other characteristics related to recent, survey and diverse set of papers are considered for a reading list, new techniques are required.

Task of Finding Similar Papers based on a Single Paper

In this subcategory, studies cover approaches for ascertaining similarity between two research papers, using both citation-based relations and textual relations. The introduction of CiteSeer digital library brought forth the CCIDF technique which was inspired by the popular TF-IDF technique (Jones, 1972). Common Citation Inverse Document Frequency (CCIDF) algorithm (Lawrence, Lee Giles, & Bollacker, 1999)

calculates similarity of a paper with all the other papers in the corpus. The algorithm factors together the citation count and co-citations of papers (White & Griffith, 1981) towards calculating the CCIDF value. Since the algorithm requires the whole corpus for processing, it is considered to be a computationally expensive method. This algorithm has been improved by combining co-references (Kessler, 1963) data in a later study (Huynh et al., 2012) where the results showed that the modified algorithm provides better results. Co-citations and co-references have been used in studies as a base model for incorporating further extensions. Such studies include (i) nested referencing in co-references used to enhance similarity calculation (Yoon, Kim, & Park, 2010) and (ii) co-citation scores calculation based on the entire network in contrast to the immediate neighbors used in the traditional method (Jeh & Widom, 2002).

Dependency based on relation type of a citation between a citing & a cited paper and graph distance in citation networks are combined to form two similarity metrics in another citation oriented study (Liang et al., 2011). These metrics are used for identifying relevant papers via depth-first-search (DFS) technique. During experimentation, the proposed approach outperformed the basic citation chaining methods, CCIDF and graph distance methods. Since the aforementioned approach was solely citation-based, it is limited to discovering papers only in citation networks. This apparent gap has been addressed in a recent study (Pan, Dai, Huang, & Chen, 2015) where citation data and textual data have been combined to form a heterogeneous graph. The graph is subsequently used in a semi-supervised learning algorithm to classify categories of similar papers. Since this is a training based approach, the model needs to be re-run whenever new papers are added. The combination of citations and textual content in formulating recommendations is adopted in another recent study

(Chakraborty et al., 2016). The recommendation technique is based on the random walk with restart (RWR) algorithm and it classifies similar papers into different facets such as alternate approaches, background and methods, thereby facilitating easier understanding of the recommended papers for researchers.

Task of Finding Similar Papers based on Multiple Papers

Mcnee (2006) proposed the use of collaborative filtering (CF) algorithms for finding similar papers. The user-item CF variant simulates the bibliographic coupling method (Kessler, 1963) while the item-item CF variant simulates the co-citation analysis (Small, 1973). Each paper in the seed set is passed one by one to the user-item matrix so that the recommendations could be generated. In the experiments, these CF variants performed better than content-based retrieval methods in finding more relevant papers. Based on a few papers-of-interest, a recent study (Küçüktunç & Saule, 2015) employed random walk algorithms for finding a diversified set of research papers. The web service *theadvisor* (Küçüktunç et al., 2013) uses the proposed technique for recommending papers. This technique is solely reliant on citation relations, so probability of finding new papers can be low.

Studies under both the subcategories are useful for researchers since the task of finding similar papers is a complex task to be performed manually. There are two shortcomings in these studies (i) majority of the studies propose recommendation techniques based on a single paper and (ii) combined textual and non-textual discovery methods are scarcely used in most studies. Solely relying on citation network might not help in finding some interesting papers.

Task of Searching Papers based on Input Text

These set of studies model the recommendations based on textual input provided by the users. The textual inputs are search keywords and dense texts such as abstracts and article drafts (along with references). As the inputs to the systems are textual in nature, most of the studies make use of content based (CB) RS algorithms for generating recommendations. The common approach is extracting references from the input draft to find relevant citations based on relational criteria that compare all the papers in the citation web (Bethard & Jurafsky, 2010; Strohman, Croft, & Jensen, 2007; Zarrinkalam & Kahani, 2013) while some studies (Zarrinkalam & Kahani, 2012) use the search keywords as input to an IR styled recommender that provides recommendations based on a novel citation metric. A combination of translation and LDA-based topic models are used in the *Refseer* system (Huang et al., 2014) which accepts bulk text as input. Phrasal concepts have been used in another study where the input draft text is used to recommend concepts extracted from candidate papers (Kim, Seo, Croft, & Smith, 2014).

These studies highlight the potential to improve current academic search systems as they are largely based on IR working mechanism. However, there is a minor drawback as they are entirely based on textual input; therefore the performance of these systems can be improved by including other input dimensions such as previously read papers of the user. Additionally, since these studies are predominantly IR based, different retrieval methods such as BM25 (Robertson & Walker, 1994) and language models (Song & Croft, 1999) could be tested out to further improve recommendations.

Disseminate Findings

In this stage of the research lifecycle, researchers have the intention of writing manuscripts for publication in journals, conferences, symposiums and workshops. For this subcategory, RS studies have provided recommendations for publication venues, citation contexts and citations based on reading list & paper draft. These studies are discussed as follows.

Publication Venues

Before writing a manuscript, researchers need to decide on the particular publication venue where the manuscript is intended to be submitted. RS techniques have been proposed this scenario for recommending publication venues. CF algorithms have been used in a study where the focus is to help novice researchers (Yang & Davison, 2012). The technique proposed in this study compares published papers from different venues to the manuscript's topic and writing style, to formulate recommendations. The co-author network has been used extensively to recommend publication venues. In (Huynh et al., 2012), the k-nearest neighborhood (kNN) algorithm is used to create a model based on the most frequent conferences attended by the co-authors. A similar approach has been used in a recent study (Beierle, Tan, & Grunert, 2016) which makes use of additional information to formulate recommendations. Specifically, four types of relations based on common publications, affiliations, similar keywords and commonly visited venues, were used for identifying similarities. The use of co-author network has been extended in another study (Chen, Xia, Jiang, Liu, & Zhang, 2015) which uses the random walk algorithm for the recommendation process.

Citation Contexts

A citation context is a location in the manuscript where the author is supposed to add a citation. Studies focusing on this concept are one of the most computationally intensive studies as a substantial part of the content of research articles are stored as computational models for making recommendations. The studies make use of content based (CB) recommender techniques as there is no scope for using CF based techniques due to lack of user ratings. Computational models such as topic models (Blei, Ng, & Jordan, 2003), language models (Song & Croft, 1999) and translation models (Berger & Lafferty, 1999) have been used in these studies as they provide the maximum compression along with accurate representation of training data. Translation models have been specifically used in (Huang et al., 2012; Lu, He, Shan, & Yan, 2011) as they are able to handle the issue of vocabulary mismatch gap between the user query and document content. The efficiency of the approaches is dependent on the comprehensiveness of training set data as the locations and corresponding citations data are recorded. The study in (He, Kifer, Pei, Mitra, & Giles, 2011) is the most sophisticated among the three, as it doesn't expect the user to mark the citation contexts in the input paper unlike other studies where the contexts have to be set by the user. The study learns the spots in previous research articles where citations are widely made so that the citation recommendation can be made on occurrence of similar patterns. The methods in these studies are heavily reliant on the quality & quantity of training data; therefore they are not applicable to systems which lack access to full text of research papers.

2.3.3.2 Passive and Implicit Information Needs (PIIN)

The PIIN category represents the set of studies where the information needs of researchers are both passive and implicit. These information needs are not stated explicitly by the researchers. Instead, the needs are ascertained based on the researcher's activities at various time periods. The subcategories under PIIN are discussed as follows.

User Footprint

User footprint refers to the user access data captured by academic search systems and digital libraries through query and access logs. Access logs are often mined in web applications to uncover users' implicit interests (Eirinaki & Vazirgiannis, 2003). Access log data is one of the most preferred source data in e-commerce based recommenders (e.g. Amazon). In the academic domain, bX recommenders¹ have already emerged as a popular service providing paper recommendations based on usage data (Vellino, 2013). The user footprint subcategory is popular among researchers with many studies (Agarwal, Haque, Liu, & Parsons, 2004, 2005; Lai & Liu, 2009; Liu et al., 2012; Vellino & Zeber, 2007; Yang, Zhang, & Feng, 2007; Zhang & Li, 2010) making use of hybrid recommendation techniques and diverse representation models such as ontology (Weng & Chang, 2008) and markov model (Chen, Chen, Guo, Yu, & Wang, 2012). There are two common approaches in the studies – (i) small subset of recently accessed articles are identified as task (Yang & Lin, 2013) or knowledge flow (Lai & Liu, 2009; Liu et al., 2012) and (ii) recommendations are based on these subsets while other studies use the entire usage history of users to build user profiles. The study of Vellino (2013) provides a comparison of usage based recommenders from this section with the citation based recommenders and makes a conclusion that these two recommenders are

¹ bX recommendation service <http://www.exlibrisgroup.com/category/bXUsageBasedServices>

complimentary in nature and should be used together to make efficient recommendations.

Researcher's Publication History

User profiles are one of the important constituents of RS as the data from the profile is used to find items (resources) with similar characteristics. Similarly, there are instances where the user would want to find papers with content similar to that of his/her earlier publications. CB recommenders dominate this section with the exception of two studies (Lee, Lee, & Kim, 2013; Sugiyama & Kan, 2013) that use user-based CF (UBCF) algorithms. Chandrasekaran et al (2008) made use of the ACM taxonomy to assign concepts to all articles in the datasets using a classifier after which these concepts are used for comparison with the user's publications to generate similar papers. In certain studies, the recommendation problem is seen as a link prediction problem to make use of citation networks along with CB recommenders to provide recommendations (Sugiyama & Kan, 2011, 2013). The vector space model (Baeza-Yates & Ribeiro-Neto, 1999) is used as the preferred user model for data storage and representation in the studies of this section. In these studies, the systems which do not allow the users to control the recommender are at a distinct disadvantage since the researcher's publication history may consist of topics in which the researcher is not interested or the interests may be outdated.

Social Network of Authors

In the social media sphere, the social network of users provides key structural data that are ideal candidates for making relevant recommendations. Twitter is a suitable example (Hannon, Bennett, & Smyth, 2010). Similarly, the network data of authors and regular users have been explored in RS to model recommendations. Majority of

these studies are based on co-author networks (Huynh & Hoang, 2012; Hwang, Wei, & Liao, 2010; Sugiyama & Kan, 2011) while a study is based on the social network of users (Chen et al., 2013). It is to be noted that co-author network data is one of the key data source for recommending publication venues in Section 7.3.3.1.4.1. Co-author networks are networks formed using the co-author data from papers. The user model is a graph in these studies due to the nature of the source data with different approaches such as closeness measures (Hwang et al., 2010), custom metrics (Huynh et al., 2012; Huynh & Hoang, 2012) and textual similarity measures such as cosine similarity (Baeza-Yates & Ribeiro-Neto, 1999) and Jaccard coefficient (Jaccard, 1912), used to make recommendations. It is observed that these recommenders are ideal for situations where the users prefer to have recommendations based on a network of selective authors. In these studies, the systems which do not allow the users to control the recommender are at a distinct disadvantage since the researcher's publication history may consist of co-authors who are one-time collaborators. Co-authors' experience level plays a major role as well. Therefore, it may not be applicable for new researchers. The number of nodes in the network is another crucial factor affecting the quality of recommendations.

Social Tags

Social Tags are annotations assigned by users for resources. Tags are one of the important constituents of the social media revolution (O'reilly, 2007). While research articles in traditional academic databases have been annotated only with the author specified keywords, user tags have been introduced in reference management websites (e.g., CiteULike², Bibsonomy³) and software (e.g., Mendeley, Endnote) to provide the

² CiteULike <http://www.citeulike.org/>

³ Bibsonomy <http://www.bibsonomy.org/>

user with the flexibility of adding his/her own tags to the indexed papers, for easier identification and retrieval. Tags have been used as a source dimension to find related papers in RS. Majority of the studies with the exception of (Joeran Beel, Langer, Genzmehr, & Nürnberger, 2013) make use of tags from the website CiteULike. In these studies, tags have been used as a means to measure user-user (Parra, 2009) and user-paper (Bogers & Bosch, 2008) similarity. The work of Guan et al (2010) is unique as it perceives the recommendation problem as a representation learning problem thereby making use of semantic space similarity (Jiang, Jia, Feng, & Zhao, 2012) to make recommendations. In (Bansal et al., 2016), recurrent neural networks has been used along with CF technique for recommending papers. In general, the solutions from these studies are applicable in environments where user tags are available and used extensively. They suffer from issues such as noise in tags, polysemic tags and unrelated tags atypical of social tagging systems (Gupta, Li, Yin, & Han, 2010). However, tag based recommenders can prove useful when combined with other input dimensions due to their ability to capture user tastes.

Periodic Recommendations from Reference Management Systems

Reference management software (e.g., EndNote, Mendeley) are used by researchers in every stage of the research lifecycle. Therefore, these systems are ideal for inclusion of recommendation modules. *Papyrus* was introduced as a novel reference management software with an inbuilt recommendation module that works on top of both content based (CB) and collaborative filtering (CF) recommender techniques (Naak et al., 2008, 2009). The multi-criteria recommendation model put forth in these studies are one of the few models introduced in the domain of SPRS. The drawback in this study is the dependency on users to provide ratings for multiple criteria pertaining to the indexed papers and user ratings are a rarity in research studies of this domain and they

have a tendency to be highly subjective. Manouselis & Verbert (2013) put forth a method for layered evaluation of multi-criteria collaborative data generated from Mendeley. The approach makes use of Multi-Attribute Utility (MAUT) (Manouselis & Costopoulou, 2007) collaborative filtering technique. However, the effectiveness of the technique is not known as the authors have not evaluated their proposed method. *Docear*⁴ is one of the latest reference management software released by the research group at University of Berkeley (Beel, Genzmehr, et al., 2013; Joeran Beel, Langer, Genzmehr, & Nürnberger, 2013). The tool has a mind map feature that helps users in better organizing their references. The in-built recommendation module in this tool is based on content based (CB) recommendation technique with all the data stored in a central server. Such systems depend on the quality and quantity of full text data available in the central server as scarcity of papers could lead to redundant recommendations. *Mendeley*'s recommendations have evolved from research papers to research profiles (Vargas et al., 2016) with the futuristic mission of positioning the tool as a “research operating system”.

2.3.3.3 Summary and Research Gaps in SPRS

Based on the survey of the published papers in the area of SPRS, certain research gaps have been identified. They are explained as follows.

Gap 1: Consolidated Framework for Contextual Dimensions

There is an inherent absence of a framework to consolidate input dimensions, filtering stages, recommender techniques/algorithms and evaluation metrics used in the previous studies. Studies have primarily concentrated on specific problems dealing with new kind of input dimensions and have paid less attention to the earlier

⁴Docear <https://www.docear.org/>

researched dimensions. When these studies are seen from the purview of contextual RS, they appear to be largely sporadic which might be considered problematic as finding relevant resources for a user's information needs requires the information system to consider multiple dimensions both from the system and the user perspectives, as highlighted by Saracevic (Saracevic, 2007a, 2007b). On the positive side, majority of these studies fall under the school of 'representational view' context which considers context to be a set of perceivable and recordable variables that can be used for recommendations. Therefore, there is an ideological unity among these studies. An effort has been taken in the direction of consolidating input dimensions with the study of Dehghani et al (2011) who have put forth a multi-layer contextual model for RS. This model captures most input dimensions such as search keywords, experience level, access log data, purpose, literacy and other minor dimensions. Future studies need to concentrate on this consolidation aspect so that there is a common ground laid for subsequent research studies. This research gap has not been selected for addressing in the current research.

Gap 2: Lack of Connectivity between Tasks

Researchers have to deal with the lack of interconnectivity across search systems in academic systems, a situation which could be mitigated with user profiles feature in RS. Mcnee in his studies (Mcnee et al., 2006a; Mcnee, 2006a; Torres, Mcnee, Abel, Konstan, & Riedl, 2004) identified seven information seeking tasks which researchers would employ during their literature search sessions. In these studies, the tasks were mapped to RS using Human Recommendation Interaction (HRI) model. However, there is no relation or connection established between these tasks. This gap needs to be addressed in order to avoid the following situations (i) repeated recommendations of same papers already read by the researcher and (ii) non-personalized generic

recommendations. This research gap is addressed by RQ2 in the current research (refer Section 1.1 in Chapter 1)

Gap 3: Lack of Relation(s) between Tasks and RS Filtering Mechanisms

Traditionally, RS algorithms have a singular task of predicting ratings for items which will be probably preferred by users. The citation recommendation setting is non-traditional in the sense that researchers do not rate papers necessarily (even though this feature is provided in contemporary reference management systems), therefore the task of formulating recommendations has to be done in an alternative fashion, as shown by Mcnee and Ekstrand (Ekstrand et al., 2010; Mcnee, 2006a). Additionally, there are different information needs for researchers looking for research papers, hence a consistent recommendation formulation process may not be applicable for different information needs (as shown by the different recommendation mechanisms used in earlier studies. In Mcnee's study, the information seeking tasks are mapped to RS by the Human Recommender Interaction (HRI). The drawback of this mapping is its subjective nature and no actual data is transferred from the tasks to the RS. A better mapping mechanism needs to be devised that meaningfully connects these two entities. This research gap is addressed by RQ2 in the current research.

Gap 4: Absence of Article-type as an Input Dimension

Based on the literature review of SPRS, very few studies have been conducted for the manuscript preparatory and writing activities. Of these activities, writing is one activity where two types of recommendation scenarios are applicable. The first popular scenario is the recommendation of articles as per the citation context (Lu et al., 2011) while the second unexplored scenario is the shortlisting service of articles from the researcher's final reading list. In this scenario, it is not just a straightforward

process whereby researchers cite all the papers that they have read. In fact, the citation of papers is performed as per the relevance of the article to the researcher's study and it is also affected by the type of the article that the researcher is writing. The article-type refers to the publication document that the researcher produces. It can be a journal paper, conference paper or a report. This dimension has been highlighted as part of the user's 'Purpose' in the multi-layer contextual model put forth in (Dehghani et al., 2011). The article-type indirectly refers to the goal of the researcher. It is to be noted that goal or purpose related dimensions have been considered for research in other research areas of recommender systems namely course recommendations (Winoto, Tang, & Mccalla, 2012) and TV guide recommendations (Setten, Veenstra, Nijholt, & Van Dijk, 2006). None of the earlier studies have considered this input dimension. Therefore, this presents an opportunity area for RS to provide recommendations in the form of a paper shortlisting service, based on the article-type. This research gap is addressed by RQ1 in the current research.

2.3.4 Summary of Part Two

Performance improvement of the algorithms from earlier SPRS studies is not a critical problem as the impact and contributions might not be considerably big. Prior SPRS studies make use of various algorithms that focus on the semantic, syntactic and structural aspects of research papers and citations. Therefore, these solutions are majorly algorithmic in nature with less modeling employed in establishing characteristics of tasks or establishing a logical connection between the tasks and the RS algorithms. This is a glaring gap as there needs to be a basic model on top of which algorithms can be employed; otherwise the base characteristics of tasks are not properly operationalized. This issue is supported by the information seeking studies where researchers (especially novices and researchers venturing in new research areas)

are seen to be in need of guided support. Therefore, the key tasks are to be identified, their characteristics outlined and a set of base features needs to be established. Subsequently, corresponding recommendation techniques can be conceptualized as part of a system so that recommendations can be provided to the users for different tasks.

Arising from research gaps identified from the literature review, there are opportunities for research and development in the following areas. The current study seeks to address the lack of connectivity/relations between tasks and RS filtering mechanisms, which are related to the goal of this research since both issues can be addressed together. Secondly, the study considers article-type as an input dimension to RS since article-type based recommendations can be conceptualized into a corresponding citation filtering task at the stage of manuscript preparation in the research lifecycle. It is to be noted that the research on creating a consolidated framework for contextual dimensions has not been considered as the solution would extend beyond the overarching goal of building an assistive system for key LR and MP tasks.

CHAPTER THREE - STUDY I: SURVEY ON INADEQUATE AND OMITTED CITATIONS (IOC) IN MANUSCRIPTS

3.1 Introduction

This chapter begins with the aims of Study I. Subsequently, the sections about the research method, survey instrument and participants are presented. The results of the statistical tests and findings are discussed next. The chapter ends with implications on how the findings help in addressing RO1 and RQ1.

3.2 Aims of the Study

Study I was designed as a predesign survey for the assistive system. Instead of explicitly requesting researchers to provide opinions about the LR search tasks for which they needed external assistance, an alternative problem-oriented approach was adopted in this study with the following rationale. In order to identify a key problem in the research lifecycle, there is a need to look at the criticality of the consequences of the problems. The inadequacies of research studies are identified at the stage where official reviewers review the manuscripts submitted to journals and conferences. Manuscript reviewers have cited issues such as missing citations, unclear research objectives and improper research framework (Bornmann, Weymuth, & Daniel, 2009; McKercher, Law, Weber, & Hsu, 2007).

From a citation analysis perspective, issues such as formal influences not being cited, biased citing and informal influences not being cited, have been identified as problems with cited papers in manuscripts (MacRoberts & MacRoberts, 1989). An improper literature review often leads to problems in subsequent stages of research since findings are not compared with the correct and sufficient number of prior

studies. Thus, inadequate and omitted citations (IOC) in the related work section of manuscripts are indicators of poor quality in research. This topic was chosen as the central topic for Study I. This study intends to investigate the issue of inadequate and omitted citations (IOC) in manuscripts and associated factors by addressing the following: questions.

Q1: What are the critical instances of IOC in research manuscripts submitted for review in journal and conferences?

Q2: Do the critical instances and reasons of IOC in research manuscripts relate with the scenarios/tasks where researchers need external assistance in finding papers?

Q3: Identify the prominent information sources that include academic databases, search engines and digital libraries, through frequency of usage.

Q4: What is the researchers' awareness level of available recommendation services for research papers?

These sub-topics surveyed in Study I is illustrated in Figure 3.1. In this figure, the instances surveyed for the sub-topics of IOC are illustrated.

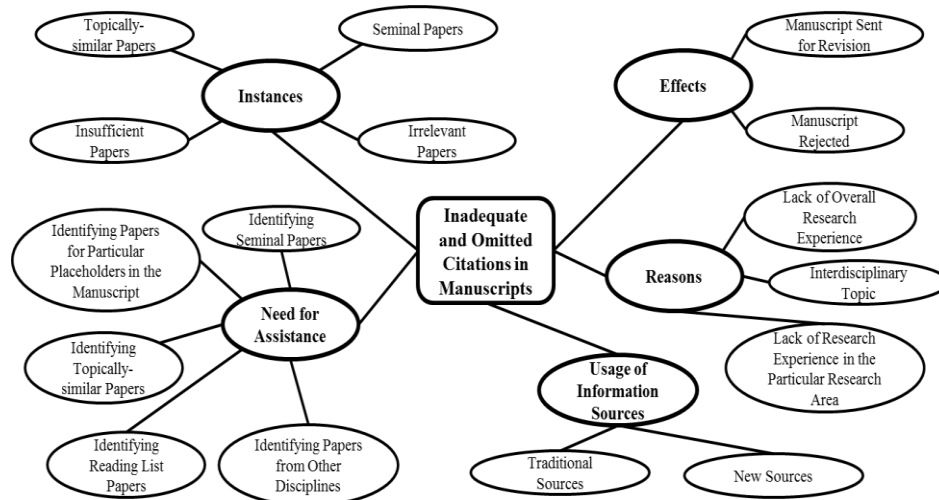


Figure 3.1. Sub-topics of inadequate and omitted citations surveyed in Study I

3.3 Method

3.3.1 Online Survey Instrument

In an earlier study conducted by MacRoberts & MacRoberts (1988), they used the interview technique to identify author motivations for citing certain papers and omitting other papers during manuscript preparation. Interviews were not considered for the study as the responses for the questions were to be measured mostly in terms of frequency of occurrence and usage. There were no open ended and exploratory questions for the participants. Thus, the aims of the study were not suited to the interview method. Secondly, the sample size is a drawback with interviews; particularly in the context of this study where there was no intent to collect qualitative data. Data for this study was hence collected through questionnaire surveys. The questionnaires were pre-tested in a pilot study. The survey flow and questions were corrected based on the feedback of the participants for the two criteria of readability and comprehensibility. Data about age group, gender, highest education level, current position, parent school, primary discipline and self-reported experience level were collected in the first section of the questionnaires. For the experience level question, the following four options were provided in the questionnaire – beginner, intermediate, advanced and expert. These levels were not explained with descriptions in the questionnaire instead the participants were expected to ascertain their level based on the distinctions between the options. The data pertaining to the study was collected in four main sections. Details about the different sections from the full questionnaire (provided to the reviewer group) are provided in Table 3.1. The questions for the segments in Table 3.1 are available in the Appendix B.

Table 3.1

Sections and the Corresponding Segment in the Online Questionnaire

Section Name	Segment	No. of Question- Items
Demographic Details	Demographic Details	9
Reviewers' Experience with Citation of Prior Literature in Manuscripts during Manuscripts Review	Instances of IOC in journal manuscripts	4
	Instances of IOC in conference manuscripts	4
	Factors affecting authors' citing behavior of prior literature in manuscripts	5
Authors' Experience with Citation of Prior Literature in Manuscripts during Manuscripts Review	Instances of IOC in journal manuscripts	4
	Instances of IOC in conference manuscripts	4
Researchers' Tasks in Literature Review and Manuscript Writing Process	Tasks where external assistance is required	5
	Characteristics of research tasks	12
Usage of Academic Information Sources	Usage of academic databases, search engines, digital libraries and SPRS	14

In the second and third sections, data about the different instances of IOC in journal and conference manuscripts were gathered from the reviewer and author

perspectives respectively. The instances (i) *Missed citing seminal papers*, (ii) *Missed citing topically relevant papers*, (iii) *Insufficient papers in the Literature Review* and (iv) *Irrelevant papers in the Literature Review* were collated from previous studies which were aimed at helping researchers in literature review (Ekstrand et al., 2010; Hurtado Martín, Schockaert, Cornelis, & Naessens, 2013; Mcnee, 2006). The instances frequency was measured using a 5-point ordinal rating scale with the following values: Never (1), Rarely (2), Sometimes (3), Very Often (4) and Always (5). The second section was hidden for participants who identified themselves as authors with no reviewing experience.

In the fourth section, questions were specifically about tasks pertaining to literature review and manuscript writing. The two segments in this section gathered data about the tasks/scenarios where researchers needed external assistance along with opinions about the characteristics of certain key tasks. The scenarios were (i) *Identifying seminal/important papers that are to be read as a part of the literature review in your research study*, (ii) *Identifying papers that are topically similar to the papers that you have already read as part of your literature review*, (iii) *Identifying papers related to your research, from disciplines other than your primary discipline*, (iv) *Identifying papers for particular placeholders in your manuscript* and (v) *Identifying papers that must be necessarily cited in your manuscripts*. These scenarios were identified from previous studies (Bae et al., 2014; Ekstrand et al., 2010; He et al., 2011; Hurtado Martín et al., 2013; Mcnee, 2006) where recommender and hybrid information systems were used to provide recommendations to researchers. The response data was measured using a 5-point ordinal rating scale with the following values: Never (1), Rarely (2), Sometimes (3), Very Often (4) and Always (5).

In the fifth section, data about the usage of information sources such as academic databases, search engines and ‘related papers’ feature in the aforementioned information sources were collected. The ‘related papers’ feature retrieves a list of similar papers for a particular paper of interest to the user. The academic information sources considered were *Google Scholar*, *Web of Science*, *ScienceDirect*, *Scopus*, *SpringerLink*, *IEEE Xplore* and *PubMed*. These sources were selected based on their popularity and applicability to multiple disciplines. The response data from the segments in this section were measured using a 5-point ordinal rating scale with the following values: Never (1), Rarely (2), Sometimes (3), Very Often (4) and Always (5).

3.3.2 Sampling

A non-probability sampling method called ‘convenience sampling’ was adopted for this study. Using convenience sampling, individuals who are willing to participate in the study (Hocking, Stacks, & McDermott, 2003) can be recruited. Therefore, the sample is easy to acquire, along with a relatively low cost (Stangor, 2014). One caveat is that the results of studies using convenience sampling are not entirely generalizable to other settings, as this sampling method does not reflect the overall population (Salkind, 2010). Notwithstanding the shortcoming, convenience sampling helps in obtaining the required sample size in a fast manner. Secondly, this sampling method helps researchers to focus on other important aspects of the study such as the design of research instruments and usage scenarios (Stangor, 2014).

3.3.3 Participants

Two groups of participants were invited to participate in the online surveys pertaining to the objectives of this study. The first group comprised of reviewers who had

officially reviewed journal or conference papers. The second group comprised of manuscript authors who had published at least one paper in a journal or conference. The first group was provided with the full questionnaire with questions that were to be answered from the perspectives of both reviewer and author. The second group was provided with a different survey web link that pointed to a questionnaire with only author-related questions. The data for the study was collected from a single location - Nanyang Technological University, Singapore between the November 2014 and January 2015. The authors sought permission from the research heads of fourteen schools in the university, for disseminating the survey invitation mail to the academic staff, research staff and the graduate research students of the schools. For schools that did not respond or declined permission to disseminate the survey participation email, the required details of the staff and students were obtained from the school websites and web directories. Prior to the survey advertisement process and data collection, the required permissions were acquired through the Institutional Review Board (IRB) of the university. Participants were paid with a cash incentive of SGD 10 for completing the study.

From a population of 1772 potential participants in the university, a total of 207 researchers (response rate of 12%) participated in the study with majority ($n = 146$ (71%)) answering from both reviewer and author perspectives since they met the qualification criteria. Schools from the engineering and natural sciences disciplines had a bigger presence with 42% and 33% of the total participants (refer Figure 3.2) since the headcount of staff and students in these schools of these disciplines are higher than humanities and business disciplines. Participants were requested to rate their experience levels. Figure 3.3 provides a chart with two related demographic variables experience level and current position. There were equal number of graduate

research students (n = 82) and research staff (n = 82) while academic staff accounted for 43 participants. Participants were required to provide their journal/conference reviewing and writing experience in years. Figure 3.4 provides a column chart with experience in number of years for three variables. The three variables are journal review experience, conference review experience and manuscript writing experience. More than half of the participants had between 0 and 5 years' experience for all the three aforementioned variables (52.3%, 54% & 57.9%).

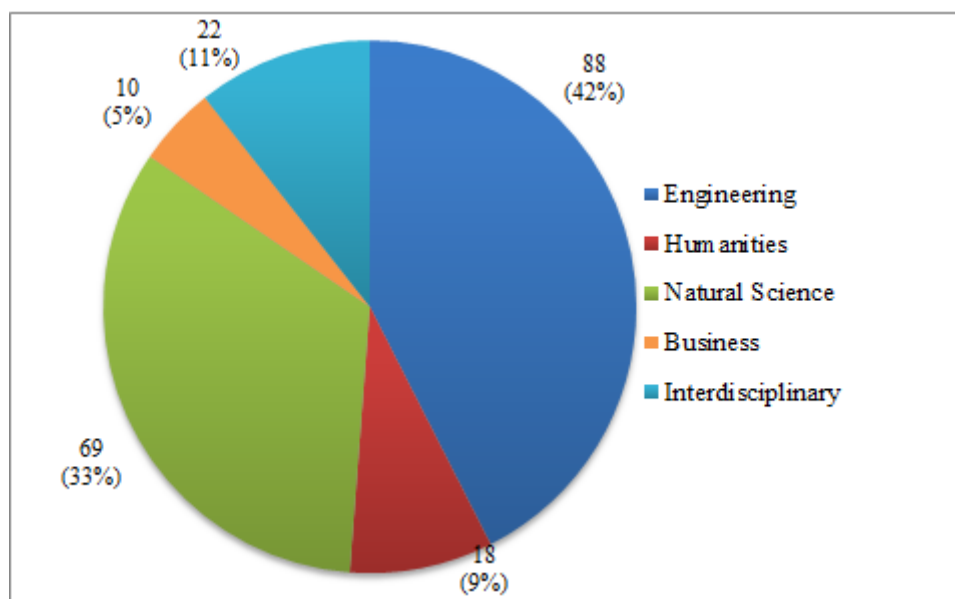


Figure 3.2. Participants data by primary discipline

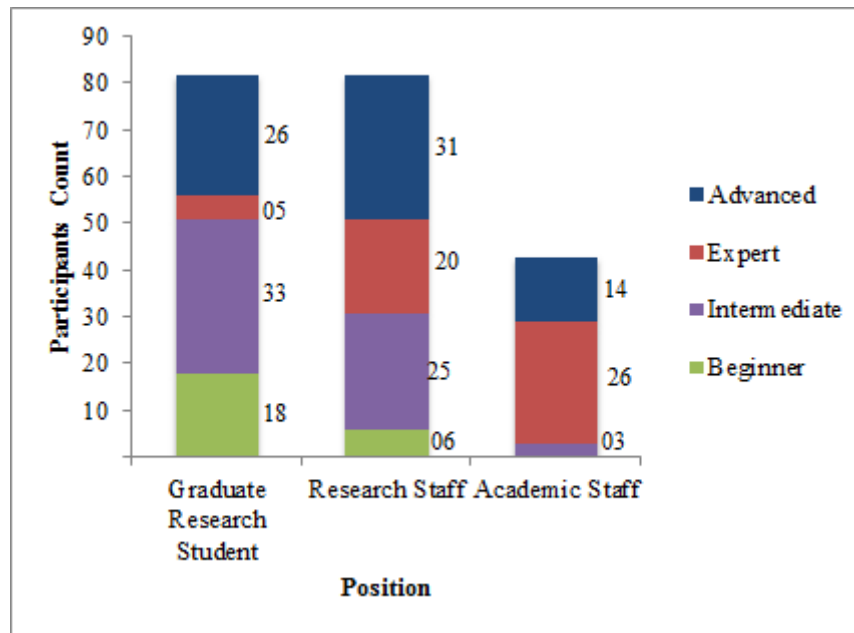


Figure 3.3. Participants data by position and experience level

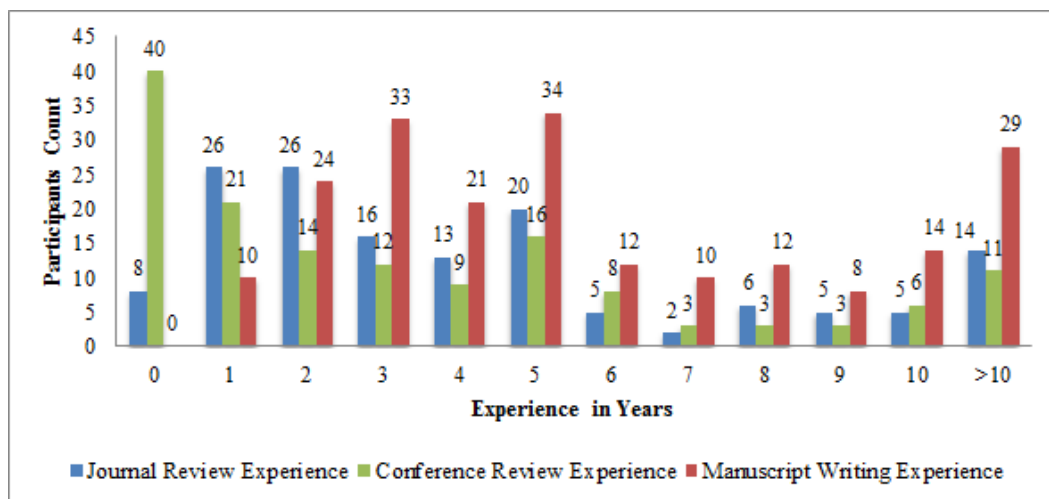


Figure 3.4. Experience of participants in journal & conference review and writing manuscripts

3.3.4 Data Analysis Procedures

Descriptive statistics were used to measure central tendency. One-sample t-test was used in the analysis to check the presence of statistically significant difference with the mean values. Statistical significance was set at $p < .05$. During the analysis, the

continuous variable writing experience was used to create a new categorical variable writing group for facilitating deeper analysis. The writing group 1 (low experience group) was allotted to observations where the participants had indicated their writing experience as less than three years. The writing group 2 (intermediate experience group) was allotted to observations where the writing experience was between four and ten years while the writing group 3 (high experience group) was allotted to observations where the experience was above 10 years. Statistical analyses were done using SPSS 21.0.

3.4 Results

3.4.1 Instances of IOC

Table 3.2 shows the results of the reviewer and author experience responses on the four instances of inadequate and omitted citations for both journal and conference manuscripts. In Table 3.3, the data is provided by the writer group values. From the reviewer perspective, the mean experience was higher than the test value of 2 (2 represents rare frequency) at statistically significant differences ($p < 0.05$) for all the four instances for journal and conference manuscripts. From the author perspective, only two instances (*Missed citing seminal papers in manuscripts*, *Missed citing topically relevant papers in manuscripts*) had mean experience higher than the test value of 2, specifically for journal manuscripts. However, all the differences in the author perspective were statistically significant.

The instance *Missed citing topically relevant papers in manuscripts* received the highest mean value for both journals ($M = 3.08$) and conferences ($M = 3.14$) from a reviewer perspective, followed by the instance *Insufficient papers in the literature review sections of manuscripts* ($M = 3.09$, $M = 3.09$). From the author perspectives,

results indicate that participants have very infrequently faced these instances in their experience. Only the instances *Missed citing seminal papers in manuscripts* ($M = 2.29$) and *Missed citing topically relevant papers in manuscripts* ($M = 2.33$) are an exception in the case of journal manuscripts.

Table 3.2

Experience of Participants on the Instances of IOC

Instances of Inadequate and omitted citations	Reviewer Perspective		Author Perspective	
	M (SD)	t	M (SD)	t
	2.83	12.718	2.29	
Missed citing seminal papers (J)	(0.770) ^a	*	(0.888) ^c	4.694*
	3.08	17.729	2.33	
Missed citing topically relevant papers (J)	(0.715) ^a	*	(0.865) ^c	5.547*
Insufficient papers in the Literature Review (J)	3.09	16.027	1.88	
	(0.797) ^a	*	(0.881) ^c	-1.972*
Irrelevant papers in the Literature Review (J)	2.78	10.163	1.41	
	(0.896) ^a	*	(0.661) ^c	-12.94*
	2.94	12.652	1.79	
Missed citing seminal papers (C)	(0.780) ^b	*	(0.848) ^c	-3.525*
	3.14	15.317	1.79	
Missed citing topically relevant papers (C)	(0.775) ^b	*	(0.859) ^c	-3.479*
Insufficient papers in the Literature Review (C)	3.09	13.155	1.56	
	(0.866) ^b	*	(0.741) ^c	-8.629*
Irrelevant papers in the Literature Review (C)	2.72		1.32	-
	(0.901) ^b	8.395*	(0.603) ^c	16.247*

Sample Size: ^an = 137, ^bn = 108, ^cn = 207, J=Journals, C=Conferences, * indicates $p < 0.05$.

Table 3.3

Mean Experience of Author Writing Groups for the Instances of IOC

Instances of Inadequate and omitted citations	Writing Group ($N = 207$)		
	1	2	3
	M (SD)	M (SD)	M (SD)
Missed citing seminal papers (J)	1.93 (0.893)	2.47 (0.879)	2.44 (0.734)
Missed citing topically relevant papers (J)	1.96 (0.878)	2.58 (0.814)	2.37 (0.757)
Insufficient papers in the Literature Review (J)	1.67 (0.86)	1.95 (0.906)	2.05 (0.815)
Irrelevant papers in the Literature Review (J)	1.43 (0.722)	1.38 (0.653)	1.42 (0.587)
Missed citing seminal papers (C)	1.72 (0.831)	1.81 (0.894)	1.86 (0.774)
Missed citing topically relevant papers (C)	1.7 (0.817)	1.87 (0.942)	1.77 (0.718)
Insufficient papers in the Literature Review (C)	1.57 (0.763)	1.53 (0.751)	1.6 (0.695)
Irrelevant papers in the Literature Review(C)	1.36 (0.644)	1.28 (0.608)	1.35 (0.529)

J=Journals, C=Conferences

3.4.2 Reasons for IOC

Table 3.4 shows the results of reviewers' opinion on the reasons for IOC. Data indicates that there is agreement in support of these reasons. The mean agreement was higher than the test value of 3 (3 represents neutral) at statistically significant differences ($p < 0.05$). The reason *Lack of Research Experience in Particular Research Area* had the highest average agreement ($M = 3.68$) than *Interdisciplinary Topic* ($M = 3.58$) and *Lack of Overall Research Experience* ($M = 3.51$).

Table 3.4

Participants Opinion on Reason for Inadequate and omitted citations (n = 146)

Reason for Inadequate and omitted citations	M (SD)	t
Lack of Research Experience in Particular Research Area	3.68 (0.813)	10.076*
Interdisciplinary Topic	3.58 (0.750)	9.382*
Lack of Overall Research Experience	3.51 (0.889)	6.892*

* indicates $p < 0.05$ **3.4.3 Effects of IOC**

Table 3.5 shows the reviewers are of the view that manuscripts are sent back for revision due to missing citations ($M = 3.10$) while they are not in agreement with case of manuscript being rejected due to that reason ($M = 2.34$). However, the difference in mean agreement with the test value of '3' is statistically significant only for the latter case at $p < 0.05$. On the other hand, Table 3.6 shows that authors have not faced the issue of their manuscript being rejected for both journal and conference manuscripts ($M = 1.32$, $M = 1.24$) with statistically significant differences with test value of 2 at $p < 0.05$.

Table 3.5

Reviewers' Opinion on Effects of Missing Citations (n = 146)

Effects of Missing Citations	M (SD)	t
Manuscript Sent Back for Revision	3.10 (0.920)	1.26
Manuscript Rejected	2.34 (0.970)	-8.272*

* indicates $p < 0.05$

Table 3.6

Authors' Experience on Effects of Inadequate and Omitted Citations (n = 207)

Effects of Missing Citations	M (SD)	t
Manuscript Rejected (J)	1.32 (0.658)	-14.778*
Manuscript Rejected (C)	1.24 (0.548)	-19.903*

J=Journals, C=Conferences, * indicates $p < 0.05$

3.4.4 Need for External Assistance in Finding Papers

Table 3.7 shows the opinions of the participants on the need for external assistance for different types of papers required during literature search sessions. The mean value was higher than the test value of 2 (2 represents rare frequency) at statistically significant differences ($p < 0.05$) for the five scenarios. *Interdisciplinary papers* are the paper-type where most external assistance is required during literature search ($M = 2.72$). The need of assistance in searching two other paper-types *Topically-similar papers* ($M = 2.6$) and *Seminal papers* ($M = 2.53$) is also visibly apparent.

Table 3.7

Scenarios for External Assistance during Literature Review & Academic Writing (n = 207)

Scenarios for Assistance during LR and Writing	M (SD)	t
Seminal Papers	2.53 (0.984)	7.696*
Topically-similar Papers	2.6 (1.009)	8.542*
Interdisciplinary Papers	2.72 (0.89)	11.71*
Citations for Placeholders	2.53 (0.954)	7.939*
Necessary Citations for inclusion in Manuscripts	2.48 (1.088)	6.327*

* indicates $p < 0.05$

3.4.5 Usage of Academic Information Sources

Table 3.8 provides the results of one-sample t-tests carried out with the variables ‘Information Sources’ and ‘Related Papers Feature’. Test values of 3 (Sometimes) and 2 (Rarely) were used for the two t-tests respectively. *Google Scholar* was the only system with a mean value above the test value ($M = 3.81$) followed by *ScienceDirect* ($M = 2.86$) (not statistically significant at $p < 0.05$) and *Web of Science* ($M = 2.79$). *PubMed* ($M = 1.9$) was the least used of all the information sources. For the variable ‘Related Papers Feature’, all the information sources had mean values above the test value and the differences were statistically significant at $p < 0.05$. The highest usage was recorded for ScienceDirect’s *Recommended Articles* ($M = 3.11$) and Google Scholar’s *Related Articles* ($M = 2.95$) while the lowest usage was for SpringerLink’s *Related Content* ($M = 2.4$).

Figure 3.5 provides the usage response data of information sources by discipline. The respondents who selected any choice other than 1 (Never) in the five-point ordinal Likert scale, were considered for this analysis. Findings indicated that the proportion of usage was similar for sources that indexed research papers from multiple disciplines. Business researchers’ use of Scopus ($n = 2$), IEEE Xplore ($n = 2$) and PubMed ($n = 4$) was minimal while Humanities researchers’ use of IEEE Xplore was almost non-existent ($n = 1$). The highest usage of the sources was from the Engineering respondents with the exception of PubMed where Natural Science researchers’ use was the highest ($n = 35$) in comparison to Engineering ($n = 29$).

Table 3.8

Usage of Academic Information Sources and Related Papers Feature

Information Source	<i>M</i> (SD)	<i>t</i>	Related Papers Feature	<i>n</i>	<i>M</i> (SD)	<i>t</i>
Google Scholar	3.81 (1.137)	10.209*	Related Articles	192	2.95 (1.152)	11.402*
Web of Science	2.79 (1.374)	-2.227*	View Related Records	152	2.6 (1.175)	6.281*
Scopus	2.19 (1.373)	-8.455*	Related Documents	108	2.83 (1.132)	7.654*
IEEE Xplore	2.08 (1.352)	-9.816*	Similar	95	2.26 (1.132)	2.266*
ScienceDirect	2.86 (1.374)	-1.467	Recommended Articles	151	3.11 (1.105)	12.375*
SpringerLink	2.44 (1.241)	-6.442*	Related Content	139	2.4 (1.054)	4.505*
PubMed	1.9 (1.25)	- 12.618*	Related Citations	88	2.6 (1.255)	4.501*

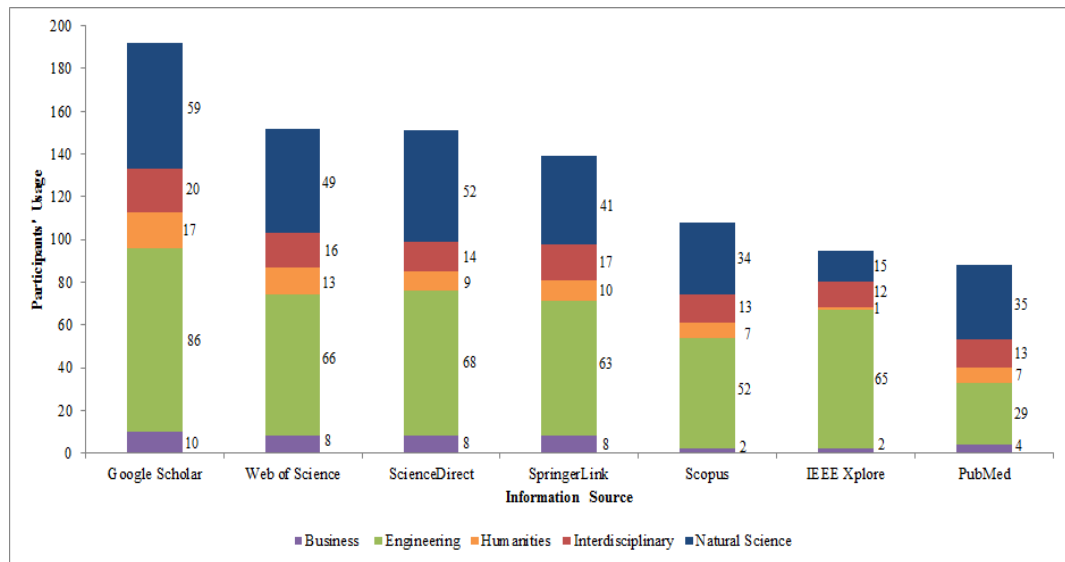


Figure 3.5. Usage of information sources by discipline

Table 3.9 provides the results of usage statistics of the available SPRS. A majority of the participants indicated that they have not used the systems ($n = 129$, 62.3%) and also haven't heard about the systems ($n = 89$, 43%). The most used system was *CSSeer* ($n = 11$, 5.3%) followed by *PubChase* ($n = 10$, 4.8%). Some participants indicated that they used other systems such as *Readcube* and *SciFinder* in addition to the four systems.

Table 3.9

Usage of Scientific Paper Recommendation Services

Choice	<i>n</i>	% of Total
Refseer	8	3.9%
Theadvisor	5	2.4%
CSSeer	11	5.3%
Pubchase	10	4.8%
Others	6	2.9%

Haven't used any of the systems	129	62.3%
Haven't heard any of the systems	89	43%

3.5 Discussion

With the exception of the two IOC instances *seminal* and *topically-similar papers* for journal manuscripts where the perceptions from the two groups were similar, the difference between the reviewers and authors experience is an interesting case. Reviewers have indicated more than a periodic occurrence of the different instances of inadequate and omitted citations, in contrast to authors' experience. In order to identify reasons behind the low frequency among authors, the data was analyzed at the writing group level. The mean values from Table 4.3 indicate that authors from the medium (group 2) and high experience (group 3) writing groups have experienced these instances more than the authors with low experience (group 1), for the first three key instances. This behavior is assumed to be due to the relatively low number of papers written by new researchers (doctoral students) and most of these papers tend to be co-written with experts (senior researchers/supervisors) (Heath, 2002). Therefore, authors might probably face these issues as they write more papers in different topics. The presence of the different instances of inadequate and omitted citations is stronger for journal papers than conference papers. This observation makes sense as conferences are used by researchers mainly for reporting interim results of their research work, in order to acquire suggestive feedback from the research community (Derntl, 2014). In summary for IOC instances, seminal and topically-similar papers are often missed by authors in their manuscripts. These findings address Q1 of Study I by identifying the two critical instances of IOC in research manuscripts submitted for review in journal and conferences.

When the reasons for IOC were identified from the literature on graduate information seeking, it was expected that the participants would agree to them as valid reasons. The findings vindicated the expectations as reviewers have opined that the three reasons - *interdisciplinary topic*, *lack of overall research experience* and *lack of research experience in the particular research area* are valid. Interdisciplinary research presents a unique kind of challenge to the researchers in terms of the information source as there are different academic databases for certain disciplines (George et al., 2006). Even though, popular academic search engines such as Google Scholar have a vast coverage of papers from most disciplines, researchers might miss out certain papers which are either not indexed or those papers may not appear in the top search results (Giustini & Boulos, 2013). General research experience is one of the key factors that differentiate researchers while doing information seeking (Karlsson et al., 2012). However, participants have indicated that the lack of research experience in the particular research area is a more prominent reason than the lack of overall research experience as there are differences in conducting research across different disciplines.

Reviewers are of the view that manuscripts are sent back to authors if the necessary citations are not included. While at the same time, both reviewers and authors have opined that manuscripts are not rejected due to inadequate and omitted citations. This observation validates prior studies as the issue does not affect the contributions or the findings of the study reported in the manuscript (Bornmann et al., 2009). Therefore, the issue of inadequate and omitted citations is not perceived as the main issue to reject a manuscript. However, it can be claimed that this deficiency will affect the overall quality of the manuscript.

Researchers have been in situations where they periodically require assistance for finding papers for all the specified tasks/scenarios. For specific search scenarios *Finding papers from other disciplines* and *Finding topically-similar papers*, researchers needed more assistance. These search tasks are generally complex tasks. Interdisciplinary research is a challenging area as it integrates methods from different disciplines (Wagner et al., 2011). Therefore, researchers would often require assistance in finding papers from other disciplines. The task of finding similar paper is an interesting one as it deals with the aspects of *scale* and *technique*.

The *scale* aspect of this task is the quantity of papers that goes into the seed set for which similar papers are found. There are various scenarios for this aspect. For example, a researcher might want to find similar papers for one seed paper while in another scenario, the researcher might want to find similar papers for a set of papers (Mcnee, 2006).

The *technique* aspect is about the method and data used in finding similar papers. Some of the commonly used techniques include (i) citation chaining (forward and backward chaining) based on a set of references (Levy & Ellis, 2006), (ii) textual-similarity techniques based on text extracted from a research paper and (iii) metadata based techniques where metadata from a paper is used to similar papers. Researchers have indicated that keeping up-to-date with latest research studies and exploring tangentially similar yet unfamiliar areas are complex tasks (Athukorala, Hoggan, Lehtio, Ruotsalo, & Jacucci, 2013). Therefore, the task is complex task where researchers would be in situations where they require additional assistance for task completion. Interestingly, the observations for the task of finding similar papers, directly relate to the observations for the corresponding inadequate citation instance. Hence, this task is highly critical for researchers. Secondly, the observations for the

task of finding interdisciplinary papers, directly relate to the corresponding second most important reason for inadequate and omitted citations. Therefore for Q2, the findings confirm that the critical instances and reasons of IOC in research manuscripts relate with the scenarios/tasks where researchers need external assistance in finding papers.

The popularity of Google Scholar among researchers across different disciplines is re-instated through the findings of this study. The finding is an indication of the constant growth in popularity for Google Scholar through the years (Cothran, 2011; Herrera, 2011). For Q3, the databases ScienceDirect and Web of Science are the other popular choices since they are two of the oldest available online academic information sources, covering multiple disciplines. Most of the databases that were included in the questionnaire are multidisciplinary with the exception of IEEE Xplore and PubMed. Therefore, these two systems were the least popular sources. The usage of all these sources is not mutually exclusive. The participants' indication of frequently using the related papers feature in the systems suggests their usability and the researcher's dependability on system-level intelligence for the task of finding similar papers. The usage of this feature is highest in the case of ScienceDirect and Google Scholar. In the case of ScienceDirect, the recommended articles are displayed to the right of the currently viewed article. A screenshot of the feature is depicted in Figure 3.6. This placement could probably attract more clicks from the users as they get to see the topical similarity in the articles placed side-by-side. The usage stats of these features indicates the reliance of researchers on the system's capability to find similar papers based on multiple factors such as topical similarity, collaborative viewership, shared citations and metadata. Google Scholar's simple and minimalist

design requires the users to scroll less and search for options. The interface can be termed as intuitive and sufficient for finding relevant information.

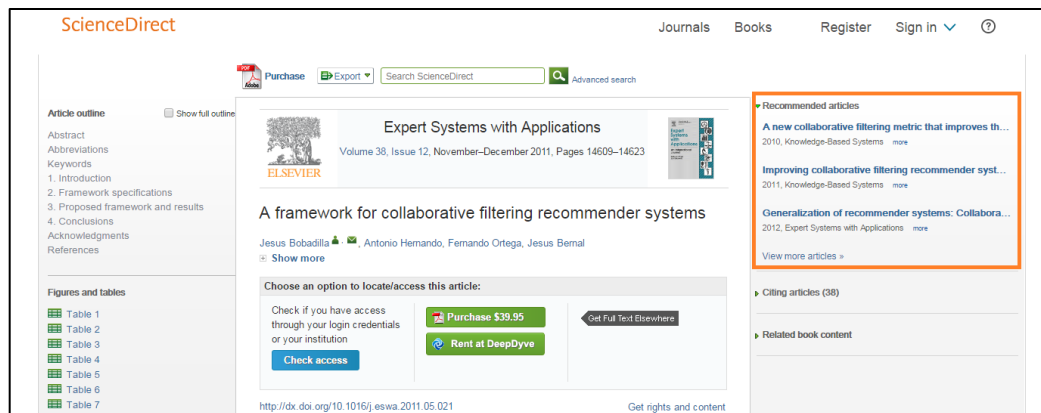


Figure 3.6. Recommended articles section in ScienceDirect

The list of SPRS considered for this study was not exhaustive. Beel & Langer (2014) have identified more such services which are currently unavailable for public access. However, the services identified as a part of this study were found to be reliable and accessible over a considerable period of time. The participants were largely unaware of these services and resultantly, the usage of these services was found to be negligible. Systems such as Readcube and Scifinder were suggested by some participants. Readcube is reference management system which provides personalized recommendations, similar to Mendeley. Scifinder is a research discovery platform that provides access to research publications from chemistry and related sciences. These two systems need not be considered as “stand-alone” recommendation services. The low popularity of recommendation services among the participants is not surprising because they are not generally promoted by university libraries. Librarians often promote library search engines and other subscription databases to students. If any additional assistance is required for graduate research students or staff, librarians assist in conducting federated search sessions for finding relevant papers. However, it

is to be seen that recommendation services can reduce the search time by recommending papers based on multiple factors. Therefore for Q4, the special purpose SPRS are largely unknown among researchers. The findings necessitate the need for promotions through librarians.

3.6 Summary

In this chapter, the findings of Study I was presented and discussed. This study has multiple implications for the assistive system. They are discussed as follows. The medium of external assistance to researchers can range from online & offline training to onsite assistance by experts and also through deployment of assistive intelligent systems. *People-oriented* interventions such as strengthening the role of librarians is a solution which is constrained by many factors such as the required knowledge levels and cultural issues (Ishimura & Bartlett, 2014). On the other hand, academic search systems could be enhanced with task-based features where search results are tailored to the specific search tasks of the researchers. Based on the findings, it is evident that researchers perceive the task of finding similar papers as a manually complex task. Secondly, seminal papers are also usually missed by manuscript authors, thereby underlining the need for a corresponding recommendation task. The need for seminal papers can be conceptualized as a reading list task where the intention is to build a reading list of papers. These papers are meant to be read at the start of LR. These identified two tasks are meant for addressing LR-related search activities. Hence, these tasks were considered for the assistive system. Interdisciplinary papers pose a challenge to researchers as it requires additional knowledge of terminologies from multiple disciplines. The retrieval/recommendation technique for the two shortlisted tasks should consider interdisciplinary papers as indicated by the explicit preference of the participants in the current study.

The selection of the third task for the assistive system is constrained by the decision to have a task meant to help researchers during manuscript preparation. In this area, citation context recommender system studies (He et al., 2011; He, Pei, Kifer, Mitra, & Giles, 2010) have concentrated on approaches to recommend papers to particular placeholders in the manuscript. A section of participants in the current study have indicated the issue of insufficient citations in their manuscripts. Therefore, assistive system could help researchers in identifying unique and important papers in their final reading list (a list comprising of all papers read or collected during LR), just before they start writing their manuscripts. This novel task is thus identified as the third task of the assistive system. The three selected tasks represent the usual flow of activity in the research lifecycle. The tasks are logically interconnected as the papers from the first task become the input to the second and third task. Therefore, paper-collection features can be introduced to help the researchers manage the papers between the tasks. In Figure 3.7, the flow of activities between the three tasks has been illustrated. In the reading list task (Task 1), an initial set of papers are first accumulated. The Task 1 papers are further expanded in Task 2 where the intent is to find similar papers, thereby further expanding the papers list. Task 2 is executed multiple times by researchers as per their need. Finally, Task 3 helps in shortlisting the whole list of papers collected from Task 1 and 3 based on the type of the article that the researcher intends to write.

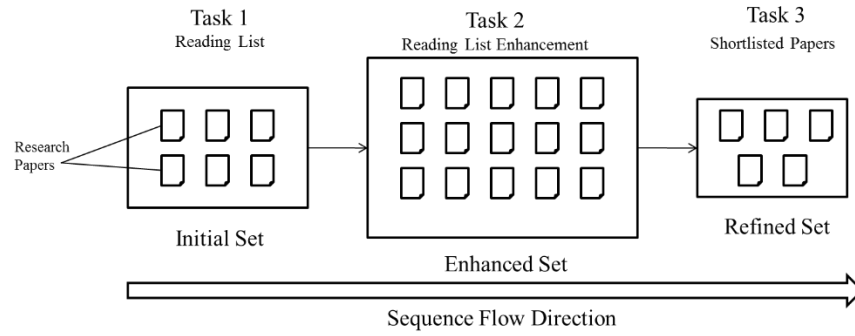


Figure 3.7. Pictorial representation of the evolution of reading list in the three tasks

The universal popularity of Google.com as a top general-purpose search engine seems to have been repeated with Google Scholar (GS) since most participants irrespective of discipline, use it frequently. This observation is validated with other studies as well (Spezi, 2016; Wu & Chen, 2014). However, GS is classified as a search engine that indexes papers from different academic databases and also from the non-academic websites while systems such as Scopus and Web of Knowledge are full-fledged databases with traditional features for search and filtering. In comparison, GS provides a basic set of advanced search options along with two sorting options. The advanced search options in GS include searching by paper title, author name(s), publication venue and publication date. GS combines a simplistic user-interface with an effective retrieval algorithm to provide fast and relevant results. If the recommendation results in the proposed system are provided in a similar user-interface as GS, it would benefit the user in quickly adjusting to the new system. Alternatively, if a new user-interface is provided to the users, separate tests are to be conducted for ascertaining the cognitive load and user convenience levels. Therefore, the option of using GS user-interface was chosen as a recommended approach for the design of assistive system. However, the display features of the system needs to be tailored as per the nature of the task (Diriye et al., 2013). For instance, a recommendation task for

finding similar papers should have a screen where the relations between the recommended papers and the seed papers are to be displayed. RQ1 was addressed in this chapter. The three tasks for the assistive system were identified through Study I.

CHAPTER FOUR - DESIGN AND DEVELOPMENT OF REC4LRW SYSTEM

4.1 Introduction

In this chapter, the technical details of the Rec4LRW system are presented. First, the seven base features that form the backbone of the recommendation techniques are described. Second, the recommendation techniques of the three tasks are introduced. In the final section, the system design aspects of the Rec4LRW system are described. The design aspects include both the task interconnectivity and informational display features in the task screens. Details regarding the performance tweaks applied to the Rec4LRW have been included in Appendix C.

4.2 Identification of Base Features

One of the drawbacks of earlier SPRS studies is the lack of an intermediate set of features for usage across RS algorithms for different tasks. Seven features that represent the characteristics of the bibliography and its relationship with the research paper were identified during this activity. The rationale for choosing bibliography is its ability in differentiating research papers. The bibliography section comprises of the references cited in the paper, indirectly representing the content of the paper. The high level characteristics of the bibliography are captured using three features: *References Count*, *Grey Literature Percentage* and *Coverage*. The next set of features was conceptualized for capturing the relations between the research paper and each reference in the bibliography. They are *Recency*, *Textual Similarity* and *Specificity*. An additional feature *Citation Count* was included as it has been traditionally used for

assessing a paper's popularity. The base features are individually described as follows.

4.2.1 References Count (RC)

The basic feature *References Count*'s value is different in the bibliographies of research papers since prior studies are mainly cited on a need basis. Also, certain publishers restrict the number of references directly or through page restrictions indirectly, thereby differentiating the count of bibliographic references across article-types. This feature *References Count* is mainly required to check for commonalities across different article-types (e.g., journal review paper, conference short paper). Data from this feature provides the potential for setting the number of the recommendations in the recommendations list provided to the user for Task 3. In context of the aforementioned scenario, the usage of this feature is novel in both citation analysis and SPRS studies.

4.2.2 Grey Literature Percentage (GLP)

Researchers usually cite references from scientific sources. However, there are situations where non-scientific references such as technical reports and websites are additionally cited. These non-scientific references that are yet to be formally published are referred to as Grey Literature (GL) (Debachere, 1995). Grey Literature is defined as “*information produced on all levels of government, academia, business and industry in electronic and print formats not controlled by commercial publishing*” (Schöpfel, Stock, Farace, & Frantzen, 2005). In a citation analysis study conducted on a sample set of articles in Google Scholar (Cesare, Luzi, & National, 2008), it was found that around 34.3% of the retrieved articles were of the GL type with technical

reports (26%) being the most used type. Articles other than conference papers, journal papers, PhD thesis and books are considered as grey literature in the current research.

GLP for a given research paper is the percentage of GL references in the bibliography. It will be used in situations where data from external sources could strengthen the recommendation list. The values calculation is performed after the identification of reference type for each reference in the bibliography of research papers. In Appendix D, the results of a citation analysis study conducted to ascertain GLP of different article-types is presented. The *GL Percentage* (GLP) on the whole for proceedings⁵ and periodicals⁶ was 17.61% and 14.48%, indicating the nature of the publications. A GL boosting technique was proposed along with a pilot study for showcasing the technique's effectiveness. The technique details are provided in Appendix D.

4.2.3 Coverage (C)

The scientific paper bibliography's ability in capturing the important references for the research topic(s) needs to be measured so that such a metric could be used to rank the papers based on their relative importance in the citation network. This metric is referred to as *Coverage*. It is generally observed that review papers have higher coverage of prior studies, in comparison to other article-types. Therefore, it was postulated that this coverage feature will distinguish the different article-types in a significant manner as it is a direct indicator of the spread of references in the bibliography.

In the current research, two novel coverage measuring techniques were proposed based on 'author-specified keywords' field in research papers. The rationale

⁵ Proceedings includes papers from conferences, workshops and symposiums

⁶ Periodicals includes papers from journals, transactions and magazines

for utilizing this field is that these keywords represent the central topics addressed in the paper. Most publications allow authors to specify a maximum of five keywords for each paper. These keywords not only provide scope for classifying the paper as detailed and broader topics but also an opportunity to identify inter-disciplinary topics. The two keywords-based coverage features are described as follows:-

4.2.3.1 Topical and Peripheral Coverage (TPC)

The *Topical and Peripheral Coverage (TPC)* is meant for utilizing all the keywords provided by the author(s) for a research paper. The measurement technique for *TPC* starts with identifying the keywords K provided for a paper P_i , followed by extracting all the papers in the corpus which have the keywords in K . This extracted set of papers becomes the base set P^k . In the next step, extraction of the references list $reflist_i$ and citations list $citelist_i$ of P_i is performed. The coverage score is measured by counting the number of papers from $reflist_i$ and $citelist_i$ that are present in P^k . Alternatively, the coverage score can be also calculated as ratio score by dividing the combined count of $reflist_i$ and $citelist_i$ by count of papers in P^k .

4.2.3.2 Topical Coverage (TC)

Since the *TPC* technique takes into account all the author-specified keywords of a research paper, it is expected to identify research papers for a diverse set of sub-topics. This approach is also suited for inter-disciplinary topics with the precondition that authors from different disciplines have used common keywords in their research papers. The second coverage feature *Topical Coverage (TC)* is conceptualized at a single research topic level. The measurement technique starts with identifying a topic T . Papers having the topic T as author-specified keyword are retrieved from the corpus to form the set P^T . In the next step, the extraction of the references list $reflist_i$ and

citations list $citelist_i$ of a particular paper P_i which has topic T as author-specified keyword is performed. The coverage score is measured by calculating the count of papers from $reflist_i$ and $citelist_i$ that are present in P^T . Alternatively, the coverage score can be also calculated as ratio score by dividing the combined count of $reflist_i$ and $citelist_i$ by count of papers in P^T . The conceptual model for TPC and TC is also provided below. The value range for both TPC and TC is from 0 to 1. A pictorial representation of the network space in these techniques is illustrated in Figure 4.1.

P_n : list of papers in the sample set of the corpus about the topic T

For each $P_i (1 \leq i \leq n)$, $reflist_i : \{P_j \mid P_i \text{ references } P_j\}$

For each $P_i (1 \leq i \leq n)$, $citlist_i : \{P_j \mid P_i \text{ is cited by } P_j\}$

K : list of all author specified keywords in P_n

P^k : list of all papers in the corpus having any k in K as author specified keyword

P^T : list of all papers in the corpus having T as author specified keyword

TPC for $P_i : \mid reflist_i \in P^k \mid + \mid citlist_i \in P^k \mid$

TC for $P_i : \mid reflist_i \in P^T \mid + \mid citlist_i \in P^T \mid$

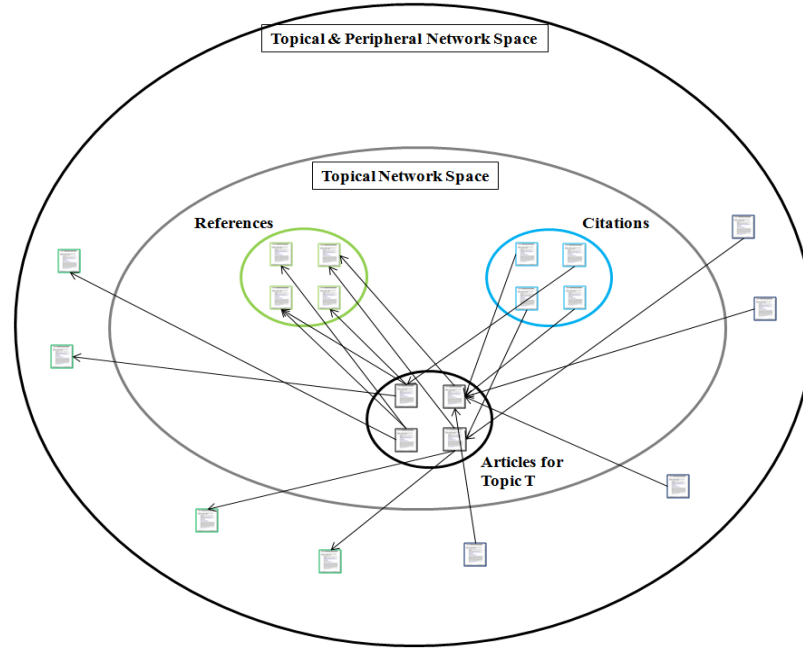


Figure 4.1. Paper network space for *TPC* and *TC* techniques

Since both these coverage values cannot be used together in a single recommendation technique, a pilot study was conducted with a single research topic ‘Information Retrieval’ to identify the best coverage value among the two. These two values were benchmarked against HITS value of a research paper where the research paper is visualized as a node in citation web/graph. HITS stands for Hyperlink-Induced Topic Search (Kleinberg, 1999). It is one of the prominent graph ranking algorithm along with PageRank (Page, Brin, Motwami, & Winograd, 1999). The detailed results of the study are presented in Appendix E. It was identified that *TPC* performs the best for diverse, popular and recent papers while HITS value performs the best for survey papers. As a conciliatory step, it was decided that the HITS value will be multiplied with the *TPC* value to form a composite value, for usage in Task 1 recommendation technique. The rationale for using the multiplication operation for this case is as follows. HITS value ranges between 0 and 1. Adding this value to *TPC* would not make a substantial difference. Therefore, the multiplication operation was

preferred since the intent was to enhance of the effects of *TPC* with *HITS*. This step ensures that the combined values could retrieve all types of papers from the corpus.

4.2.4 Recency (RE)

The temporal aspect of citations is ignored in most studies, on the basis that it does not affect the citing behavior. An earlier study indicate that new publications take an average of two years to be cited (Garfield, 2003). Therefore, the temporal data is required for recommending recent papers. Recency refers to the characteristic that shows how recent the referenced papers are in the bibliographies of the papers. It is calculated by finding difference in years between the publication date of the parent paper and each reference in the bibliography. A similar feature has been used in an earlier study (Bethard & Jurafsky, 2010). Recency values were calculated by subtracting the publication date of the parent paper and the reference, at the year level.

4.2.5 Textual Similarity (TS)

Researchers find relevant papers for their literature review by using appropriate keywords. The search keywords are topically related to their search requirement and information need. It is observed that the title of the paper is textually related to the title of the bibliographic references in most occasions. Therefore, it is imperative to calculate the similarity between the parent paper and references in the bibliography. *Textual Similarity* is the feature for calculating the keyword-based similarity of the title metadata of the parent paper and the references in the bibliography of the parent paper. A similar feature has been used in (Bethard & Jurafsky, 2010). Cosine similarity is the traditional similarity measure used in IR systems. The measure suffers in situations where the textual content is minimal.

For the current research, the bigram-based dice coefficient (Brew & McKelvie, 1996) was used for calculating the similarity between the paper title and the reference/citation title. Dice coefficient performs better than other methods such as soundex (Holmes & McCabe, 2002) and edit distance (Kukich, 1992). The formula for *Textual Similarity (TS)* between two strings $S1$ and $S2$ is given below in Equation 4.1.

$$TS(S1, S2) = \frac{2 \times |pairs(S1) \cap pairs(S2)|}{|pairs(S1)| + |pairs(S2)|} \quad (4.1)$$

Where $pairs(X)$ is a function that generates the pairs of adjacent letters (characters) from the string.

4.2.6 Specificity (S)

The previous feature *Textual Similarity* measures the similarity using the text from the title and references/citations. It is disadvantageous for certain cases where the references are related to the parent paper even though the title may contain dissimilar keywords (e.g. data mining and machine learning). A novel feature known as *Specificity* was introduced to address this issue. It refers to the nature of the references in being very specific or otherwise to the topic(s) of the parent paper. For the current research, the feature was meant to identify the similarity between a research paper and its references/citations based on commonality in the metadata fields: author-specified keywords, primary category and secondary category. The formula for *Specificity (SP)* between a paper P and a reference/citation F is given below in Equation 4.2.

$$SP(P, F) = CK(P, F) + CP(P, F) + CS(P, F) \quad (4.2)$$

Where F is a citation or reference of paper P . CK is a function that counts the number of shared author-specified keywords between P and F . CP is a function that

counts the number of shared primary categories between P and F . CS is a function that counts the number of shared second categories between P and F .

4.2.7 Citation Count (CC)

A common behavior among researchers is to cite popular papers since they are widely accepted in the research community. There is a proclivity in citing papers based on its citation count. This behavior is common while writing journal papers when compared to conference papers where the most recent works are cited even though they may not have high citation counts. The feature *Citation Count* is included as it remains one of the foremost metrics to rank popular papers.

The relations between the research paper and citations/references are depicted in Figure 4.2. In the figure, the external container is the parent paper. The features except *Citation Count* connect the parent paper to the bibliography which in turn consists of references from 1 to N where N is the total number of references. The features are placed in dotted rectangles. It can be seen that the features *References Count*, *Grey Literature Percentage* and *Coverage* are at the bibliography level while other four features are at the individual reference item level for each bibliography. The features values are to be measured for all the research papers in the corpus. These features are not the only exclusive set of features for capturing the relationship. Other features can be devised based on specific requirements. Figure 4.3 provides an illustration on the sequence of activities employed for measuring the feature values for research papers datasets.

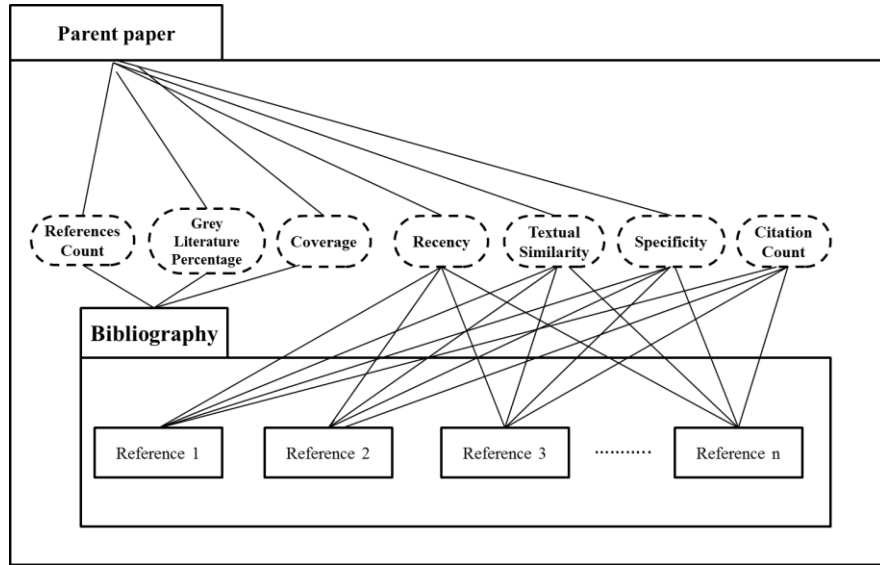


Figure 4.2. Depiction of the seven features in the research paper context

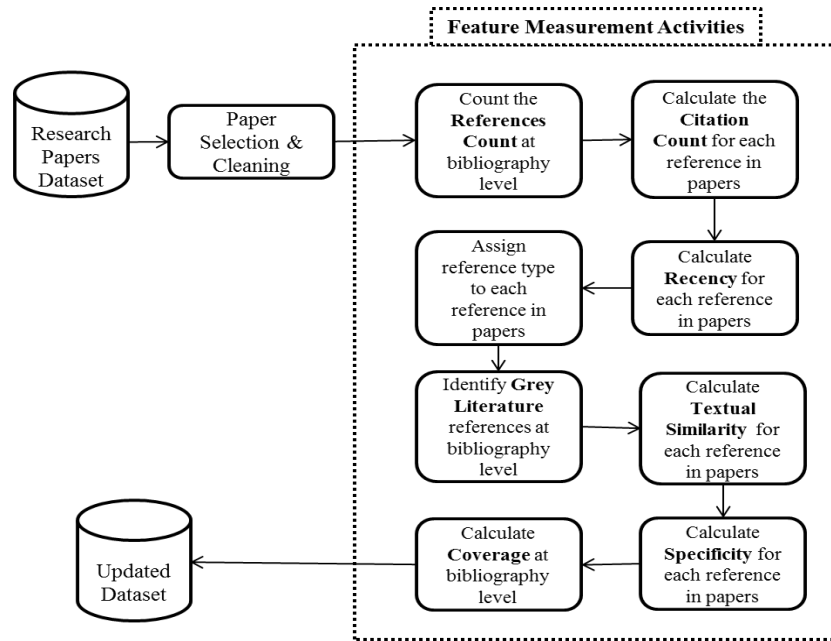


Figure 4.3. Sequence of activities in feature measurement exercise for the shortlisted papers

4.3 Recommendation Techniques for the Three Tasks

The three task objectives are re-iterated again. They are (i) building an initial reading list of research papers, (ii) finding similar papers based on set of papers, and (iii) shortlisting articles from reading list for inclusion in manuscript. Before the actual

recommendation design process, the characteristics of the tasks were analyzed for better understanding the basic characteristics. For this purpose, prior SPRS studies for the three tasks were collectively analyzed using a facet classification exercise.

4.3.1 Ascertaining the Characteristics of the Three Tasks

A faceted task classification scheme put forth in (Li & Belkin, 2008) was used for the analysis purpose. This scheme was developed in order to better understand work tasks and information search tasks of users. Additionally, it also helps in identifying relationships between the tasks. The scheme comprises of two facet types (i) Generic and (ii) Common for which there are eight facets and corresponding sub-facets. The possible values for the facets and sub-facets from the schemes are listed in Appendix F. Table 4.1 provides the facet values that were manually identified for the three tasks in the current research study. This facet classification exercise helped in better understanding the tasks at a detailed level so that good quality recommendations could be subsequently conceptualized for the tasks.

Table 4.1

Facet Values for the Three Tasks Identified for the Assistive System

Facet Type	Facet	Task 1	Task 2	Task 3
Generic	Source of Task	Internal	Internal	Internal
	Task Doer	Individual	Individual	Individual
	Frequency (Time)	Unique	Intermittent	Unique
	Length (Time)	Short-term	Long-term	Short-term
	Stage (Time)	Beginning	Middle	Final

Facet Type	Facet	Task 1	Task 2	Task 3
Common	Product	Intellectual	Intellectual/Decision	Physical
	Process	One-time task	Multi-time task	One-time task
	Quality (Goal)	Specific goal	Mixed goal	Specific goal
	Quantity (Goal)	Single goal	Multi goal	Single goal
	Objective Task Complexity	High Complexity	High Complexity	Moderate
	Interdependence	Moderate	Moderate	Moderate
	Salience of Task	High Salience	High Salience	Moderate
	Urgency	Immediate	Immediate	Moderate
	Difficulty	High Difficulty	High Difficulty	Moderate
	Subjective Task Complexity	High Complexity	High Complexity	Moderate
	Knowledge of Task Topic	Moderate	Moderate	Low knowledge
	Knowledge of Task Procedure	High knowledge	High knowledge	Moderate

Note: Task 1 refers to building an initial reading list of research papers, Task 2 refers to finding similar papers based on set of papers and Task 3 refers to shortlisting articles from reading list for inclusion in manuscript

The first two tasks can be deemed to be highly complex while Task 3 is moderately complex in terms of task execution skills and knowledge required to perform the tasks as the researcher knows which articles are the most relevant to be cited, to a decent extent. Tasks 1 and 3 are one-time tasks while the Task 2 is intermittent as the need for more papers always exists for researchers. Regarding the stage of requirement, the tasks are sequential in nature as expected since they form a part of publishing lifecycle although the execution time is lot higher for Task 2 when compared to other two tasks. The identified facet values helped in guiding the design of the recommendation models for the three tasks to an extent.

4.3.2 Task Redesign

The business-process approach of breaking up a task into input, processing and output components (Sethi & King, 1998) was considered practical and adequate for redesigning a task. These three components are the most elemental components of any process vis-a-vis task. The results of the task redesign exercise conducted for the three tasks are listed component-wise in Table 4.2. The values for the three components in the current research column in Table 4.2 have been added based on the requirements identified for each task. These requirements are listed in subsequent sub-sections for each task.

Table 4.2

Task redesign components

Task	Component	Earlier Studies	Current Research
Task 1 - Building an initial reading list	Input	Research Topics or Seed Papers	Research Topics
		Usage of citation network and paper content	Usage of citation network through author-specified keywords
	Processing	Seminal or Popular papers	Recent, Diverse, Survey and Popular papers
	Output	Single paper	Multiple papers
Task 2 - Finding similar papers based on a set of papers	Input	Usage of either citation network or paper content	Usage of both citation network and paper content
	Processing	Similar papers	Similar papers
Task 3 - Shortlisting papers from the final reading list for inclusion in manuscript based on article-type choice	Output	Not Applicable	Full list of papers
	Input	Not Applicable	Usage of citation network
	Processing	Not Applicable	Unique and highly cited papers

4.3.3 Task 1 - Building an Initial Reading List of Research Papers

4.3.3.1 Prerequisites

A reading list is colloquially defined as “*a list of sources (recommended by a teacher or university lecturer) which provide additional or background information on a subject being studied*” (Collins, 2016). There have been no formal definitions set for this collection of papers in the context of literature review (LR). Papers considered as seminal or classical in a particular research area, have constituted the reading list in prior studies (Bae et al., 2014; Ekstrand et al., 2010; Jardine, 2014; Wang et al., 2010). The common aspect of these papers is the relatively high citation counts. Even though, it is necessary for a researcher to read such important papers, such papers may not provide the overall outlook of the research area. Conversely, a researcher is expected to get a holistic understanding of the research area at the start of LR. The scientific information seeking model (Ellis et al., 1993) highlights the exploratory nature of search tasks during the initial stages of information seeking. Researchers make use of the initial set of papers in this pre-focus stage so that they could zoom into the sub-topics. These sub-topics are used in subsequent directed search sessions (for instance, Task 2) where the exact problem is formulated. Therefore, a variety of papers are expected by researchers in this first LR task. Accordingly, four requirements were proposed for this reading list which is meant to be utilized at the start of LR.

Requirement 1 (R1): The reading list should contain popular papers. As noted earlier, popular papers are the only constituents of reading lists in earlier studies. These papers have a seminal status in the particular research area. Generally, these papers have very high citation counts, indicating their popularity. Therefore, an adequate quantity of

popular papers is required for a reading list. Earlier studies (Bae et al., 2014; Ekstrand et al., 2010; Wang et al., 2010) have mainly focused on this requirement.

Requirement 2 (R2): The reading list should contain survey papers. Literature survey or review papers provide an overview of the existing research performed in a research area. Researchers generally read survey papers as a starting point in their literature review. These papers provide synthesis of prior studies along with the problem areas and research opportunities. Woodward (1977) notes that review/survey papers are widely accepted information resources for researchers. He identifies seven functions for review papers. They are (i) informed notification of the published literature, (ii) current awareness of the related fields, (iii) back-up to other literature searching, (iv) searching for alternative techniques, (v) initial orientation in a new field, (vi) teaching aids and (vii) feedback. These functions clearly highlight the fundamental importance of survey/review papers. Ideally, a reading list should contain at least one or two survey papers.

Requirement 3 (R3): The reading list should contain recent papers. Prior studies and current search systems give lesser preference to recent papers. Recently published papers provide information about the latest research performed in a research area. Additionally, recent papers potentially cite important papers. Researchers can look at such bibliographic references to discover other interesting papers. In the scientific information seeking model (Ellis & Haugan, 1997), Ellis alludes to the search for papers that help in maintaining awareness of research topics during the *Browsing* stage. Hence, the presence of recent papers in a reading list is important.

Requirement 4 (R4): A reading list should contain papers from sub-topics of the main research area. Research areas comprise of sub-topics which span out to become self-

contained research areas in themselves over a period of time. Additionally, there are interdisciplinary research areas which are subject to different research methods from corresponding disciplines. Therefore, diversity is a necessary characteristic for a reading list. A diversified reading list is meant to provide a bird's eye of the whole research landscape, particularly for broad research topics. Diversity in citation recommendations has been researched in earlier SPRS studies (Küçüktunç et al., 2015). However, diversity has not been explored in the context of this task.

Based on the above requirements, a reading list is defined as “*a list with an agreeable mix of popular, recent and survey papers covering diverse sub-topics in the particular research area*”. In the current research, the number of papers for each requirement is not fixed as a separate study needs to be conducted with researchers across disciplines for ascertaining the paper count for each requirement.

4.3.3.2 Author-specified Keywords based Retrieval (AKR) Technique

The retrieval technique for this task was operationalized by keeping *TPC* value as the main ranking entity. Since the technique is conceptually based on author-specified keywords, it is called AKR technique where AKR stands for *Author-specified Keywords based Retrieval*. The input to this task is the research topic. Earlier studies have used different types of input into their respective techniques. Ekstrand et al. (2010) used a set of seed papers as input. Wang, Zhai, Hu, & Chen (2010) trained their system with a set of papers for classification purposes. Bae et al. (2014) used a set of seminal papers as input. For this task in the current research, the naturalistic scenario of using research topic as the input was considered. The research topic is expressed as search keywords by users in academic search systems. The AKR technique is split into two stages. The process flow of the AKR technique is depicted in Figure 4.4. The

composite rank in Stage 2 involves the usage of references count, citation count and *TPC* value. Figure 4.4 illustrates the flow of activities in the AKR technique, Detailed information about the filtering process is presented in the below section.

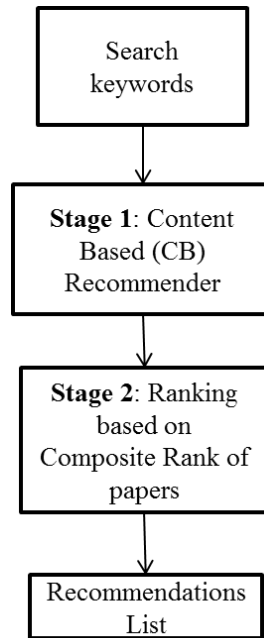


Figure 4.4. Process flow in the AKR technique

The AKR technique can be best described in two stages. These two stages help in delineating the two important functions of filtering and ranking.

Stage 1 - Content-based Filtering

The objective of this stage is to construct a representative list of papers related to the input research topic. Three metadata fields article title, article abstract and author-specified keywords are used for the text matching. These fields are merged to form a single field so that the similarity matching efficiency could be maximized. The Okapi BM25 similarity score (K Sparck Jones, Walker, & Robertson, 2000) is used for the similarity matching. The top 200 matching papers are retrieved to form set *S*.

Stage 2 - Ranking the Final List of Papers

The 200 papers from S are ranked based on ranking scheme which is majorly based on the TPC value. The objective of this stage is to shortlist the top 20 papers. The count of top papers is set as 20 since users rarely go beyond the top 2 or 3 search results pages (Van Deursen & Van Dijk, 2009). A composite rank CR_P is used for ranking papers from S . In order to give importance to citation and bibliographic reference counts albeit to a lesser level, the composite rank CR_P is based on three values – *citation count*, *reference count* and TPC . Weights are used to set the importance of these values. The sum of these weights should add up to 1. The inclusion of these weights is to vary the preference given to the three values. Weights are essential because these values capture the importance of the parent paper in the citation network at relative and global levels respectively. TPC captures the relative importance of a paper since it is based on the author-specific keywords while *citation count* and *references count* captures the overall (global) importance of the paper in the citation network. These values are potentially affected by the size of the corpus and also metadata availability in the papers. Hence, the weights serve as a mechanism to alter the preference for these values. All the three values are normalized before the CR_P is calculated. The formula for CR_P is provided in the Equation 4.3 below.

$$CR_P = \left(\left(\frac{CC_P - \min_S CC}{\max_S CC - \min_S CC} \right) * W_{CC} \right) + \left(\left(\frac{RC_P - \min_S RC}{\max_S RC - \min_S RC} \right) * W_{RC} \right) + \left(\left(\frac{TPC_P - \min_S TPC}{\max_S TPC - \min_S TPC} \right) * W_{TPC} \right) \quad (4.3)$$

Where CC_P is the citation count of paper P from set S . $\min CC$ is the minimum *citation count* value from S . $\max CC$ is the maximum *citation count* value from S . W_{CC} is the importance weight for *citation count*. RC_P is the *references count* of paper P from S . $\min RC$ is the minimum *references count* value from S . $\max RC$ is the

maximum *references count* value from S . W_{RC} is importance weight for reference count. TPC_P is the TPC value of paper P from S . $\min TPC$ is the minimum TPC value from S . $\max TPC$ is the maximum TPC value from S . W_{CO} is the importance weight for TPC value.

It is to be noted that the Topical and Peripheral Coverage (TPC) value is measured by utilizing all the author-specified keywords provided in a research paper. Since the value calculation is based on author-specified keywords, both broad and narrow topics are covered in the base set P_k (refer Section 4.2.3.1), thereby directly addressing requirement R4. On top of this setup, papers with either high number of bibliographic references or citations tend to get higher TPC scores. Hence, the other three requirements are addressed. R1 papers have high citation counts while R2 and R3 papers are expected to have high number of bibliographic references. However, R3 recent papers with low number of references will get low scores.

Since the AKR technique generates a composite rank for a given paper, this numeric value can be combined with other metrics on a need basis. For instance, graph ranking algorithms PageRank and HITS generate similar numeric values based on the structural properties of a paper in the citation network. These values could be combined with the composite rank of the AKR technique to the boost candidate papers in the final recommendation list. This approach for enhancing the AKR technique was identified when a case study experiment was conducted to ascertain the effectiveness of the TC and TPC feature values (described in Appendix E). The pseudo-code for the AKR technique is provided below.

1. Input the research topic T
2. Set variable *recommendations_counter* to N

3. Initialize the lists *base_set*, *recommendations_set*
4. Run the BM25 Similarity algorithm to retrieve the top 200 matching papers for topic *T*
5. Populate the list *base_set* with the papers from Step 2
6. Retrieve the *TPC* values for the papers in *base_set*
7. Retrieve the *CC* values for the papers in *base_set*
8. Retrieve the *RC* values for the papers in *base_set*
9. Calculate the composite rank *CR_p* for papers in *base_set* using the *TPC*, *CC* and *RC* values
10. Set variable *temp_counter* to zero
11. Set variable *M* to zero
12. Rank the papers in *base_set* based on *CR_p* values
13. While *temp_counter* is equal to *recommendations_counter*
 - Set *M* to the sum of *M* and 1
 - Populate the list *recommendations_set* with the paper in *base_set* at rank *M*
 - Set *temp_counter* to the sum of *temp_counter* and 1
14. End While
15. Print the *recommendations_set*

4.3.4 Task 2 - Finding Similar Papers based on Set of Papers

4.3.4.1 Prerequisites

The task of finding similar papers based on multiple papers is intrinsically different from finding similar papers based a single paper. In contrast to previous studies, the recommendation technique should consider all the seed papers (input paper) together while formulating recommendations. It is important to re-introduce the general

approaches followed by researchers with the aid of academic search systems, databases and citation indices. The proposed technique is meant to operationalize these basic approaches so that the recommendation of papers roughly simulates the manual process followed by researchers. There are three fundamental approaches. In the *chaining* approach, researchers generally employ the *Intellectual Structure* method (White & Griffith, 1981) that involves the dual steps of backward and forward chaining in the underlying citation network of a research paper. In backward and forward chaining, bibliographic references and citations are mined respectively for finding similar papers. Complex chaining methods such as bibliographic coupling (Kessler, 1963) and co-citation analysis (White, 1990) are used to find relevant papers based on co-references and co-citations respectively. In the *metadata hyperlinking* approach, researchers generally try to follow the publication trails of certain authors to find recent or even old papers written on the same topic. The same behavior is repeated for journals, conferences and author-specified keywords. In most of the academic systems, metadata fields of research papers are displayed as hyperlinks. These hyperlinks in principle, facilitate the option of using *follow-your-nose* method (Hausenblas, 2009) for finding relevant papers by following hyperlinks between papers. In the *extended topical searching* approach, researchers make use of certain specific terms from the seed paper(s) for further searching. These terms could be extracted from the title, author-specified keywords, abstract and the full text of a research paper. Using this approach, researchers become cognizant of the different sub-topics in the particular research area. Along with these three approaches, three natural requirements set for this task. They are as follows.

Requirement 1 (R1): The researcher may add papers from different sub-topics of a particular research area, into the seed basket. For instance, a seed basket of five

Information Retrieval (IR) papers could comprise of two papers on ranking and three papers on query representation.

Requirement 2 (R2): The researcher may add papers from different research areas into the seed basket. This scenario occurs for interdisciplinary and multidisciplinary research.

Requirement 3 (R3): When searching for similar papers, the relevance criteria for the researcher might vary as the aim is to find papers that are topically-similar albeit the new papers are not expected to be reporting very similar outcomes as the seed basket papers. Researchers searching papers for writing a survey or meta-analysis paper are exceptions to this scenario.

4.3.4.2 Integrated Discovery of Similar Papers (IDSP) Technique

The steps in the technique were put forth with the aim of simulating the three aforementioned fundamental approaches used by researchers. Three methods are used to find similar papers based on the input set of papers. The methods are classified under two modules. The process flow of the AKR technique is depicted in Figure 4.5. There are two stages in the technique. In the first stage, similar papers are discovered using the three methods. In the second stage, the papers from the three methods are merged to form the final output.

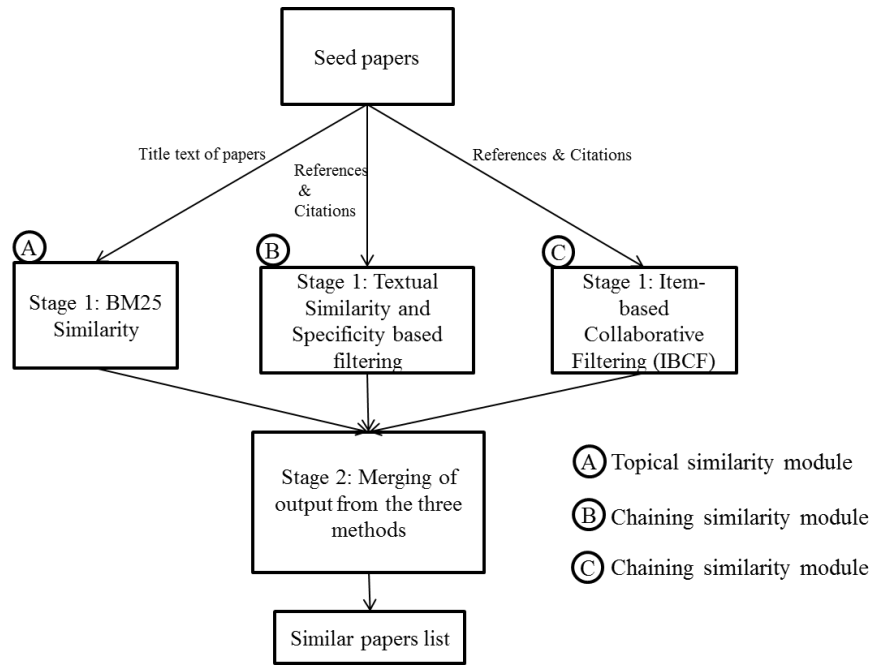


Figure 4.5. Process flow in the IDSP technique

Stage 1- Paper Discovery Modules

Topical Similarity Module

This module is meant to simulate the *extended topical searching* approach. For the current research, the ‘article title’ field is the singular field considered for computing similarity. Text from the title fields of all the seed set papers is concatenated to form a single string. This string becomes the input query. The Okapi BM25 similarity score (K Sparck Jones et al., 2000) is used for the similarity matching of the query with the documents in the corpus. The BM25 method is used since it offers better performance than other retrieval models (Speriosu & Tashiro, 2006). The top 200 matching papers are retrieved to form Set A.

Chaining Similarity Module

Collaborative Filtering Method.

In earlier studies (Ekstrand et al., 2010; Mcnee, 2006), the collaborative filtering (CF) algorithm has been found to perform better than content-based RS algorithms. The item-based CF variant (IBCF) is selected as it has provided better results than user-based CF variant (UBCF). In the user-item matrix of the IBCF algorithm, the user rows are the research papers while the item columns are occupied by the references and citations of the corresponding research papers. The rating is set to 1 between a user and item (unary item space) since there are no ratings between papers and citations. The value 1 is set if a paper cites the reference. The pictorial representation of the matrix is presented in Figure 4.6. Five recommendations are retrieved for each seed basket paper to form set *B*.

	Reference 1	Reference 2	Reference 3	.	.	Reference N
Article 1	1	1				1
Article 2			1			
Article 3	1		1			1
.						
.						
.						
.						
Article N	1					1

Figure 4.6. User-item matrix in IBCF algorithm

Feature-based Filtering

While traditional chaining methods are based on co-occurrences, the relation between a paper and its citations/references can be inferred through further analysis. In feature-based filtering method, both textual and non-textual relations were measured. The textual and non-textual relations are measured using the features *Textual Similarity (TS)* and *Specificity (S)*. As mentioned earlier in Sections 4.2.5 and 4.2.6, the bi-gram

based dice coefficient is used for *TS*. Specificity *S* is summation of the total count of common author-specified keywords, primary and secondary categories between the citation and the parent paper. For each seed paper's references and citations, *TS* and *S* values are retrieved and ranked based on descending order of their values. Through this manner, the *Textual Similarity* and *Specificity* values of research papers are combined to form set *C*.

Stage 2- Merging of Outputs from the Two Modules

The top 20 papers from each of the three sets *A*, *B* and *C* from the two modules were merged to form set *D*. The final recommendation list *L* from set *D* is prepared in the following manner. First, the papers that are already present in the user's initial reading list (Task 1 recommendations) are excluded. Second, the papers that are present in all the three sets *A*, *B* and *C* are retrieved from *D* to the final recommendation list *L*. The remaining papers in *D* are ranked based on the descending order of *citation count* of the papers. The remaining papers for *L* are retrieved based on their respective positions in *D*. It is to be noted that *citation count* metric has been used for ranking in this final step because it indicates popularity of papers. The individual scores of the papers from sets *A*, *B* and *C* have not been used in the final step since usage of these scores creates a bias favoring the corresponding filtering method. The count of recommended papers in *L* can be adjusted as per requirement. The pseudo-code for the IDSP technique is provided below.

1. Initialize the lists *seed_set*, *refcit_set*, *base_set1*, *base_set2*, *base_set3*,
temp_set, *recommendations_set*
2. Input the seed papers into the list *seed_set*
3. Set variable *recommendations_counter* to *N*
4. Retrieve the title text of the papers in the *seed_set*

5. Retrieve the bibliographic references R of the papers in the $seed_set$
6. Retrieve the citations C of the papers in the $seed_set$
7. Populate the list $refcit_set$ with R and C
8. Run the BM25 Similarity algorithm to retrieve the top 200 matching papers for concatenated titles of papers in $seed_set$
9. Populate the list $base_set1$ with the papers from Step 8
10. Retrieve the TS values for the papers in $seed_set$ and $refcit_set$
11. Retrieve the S values for the papers in $seed_set$ and $refcit_set$
12. Set the variable $feature_score$ with the sum of TS and S values for the papers in $seed_set$ and $refcit_set$
13. Populate the list $base_set2$ with the top 200 ranked papers from Step 12 based on $feature_score$ values
14. Populate the user-item matrix UI with papers from $seed_set$ and $refcit_set$
15. Run the in IBCF algorithm with UI matrix to generate recommendations
16. For each paper P in $seed_set$
 17. Recommend five papers from the IBCF algorithm
 18. Populate the list $base_set3$ with papers from Step 17
19. End For
20. Populate the list $temp_set$ by merging papers from the lists $base_set1$, $base_set2$, $base_set3$
21. Set variable $temp_counter$ to zero
22. While $temp_counter$ is equal to $recommendations_counter$
 - For each paper P in $temp_set$
 - If the occurrence count of P is greater than two
 - Populate the list $recommendations_set$ with the paper P

Set *temp_counter* to the sum of *temp_counter* and 1

Remove paper *P* from *temp_set*

End If

If the occurrence count of *P* is greater than one

Populate the list *recommendations_set* with the paper *P*

Set *temp_counter* to the sum of *temp_counter* and 1

Remove paper *P* from *temp_set*

End If

End For

Retrieve the citations count *CC* of the remaining papers in *temp_set*

Rank the remaining papers in *temp_set* based on the *CC* values

Populate the list *recommendations_set* with the remaining papers in
temp_set

23. End While

24. Print the list *recommendations_set*

4.3.5 Task 3 - Shortlisting Articles from Reading List for Inclusion in Manuscript

4.3.5.1 Prerequisites

The objective of this task is to help researchers in identifying important and unique papers from the final reading list. These papers are to be considered as potential candidates for citation in the manuscript. The article-type is used as one of the inputs to this task with the aim of differentiating the recommendations. At this stage, the researcher has a prospective publication venue article-type in mind. Although the number of papers accumulated during LR varies between researchers, it can be safely

assumed that the paper count would comfortably be at least 50. Even though the subject matter and the contributions of these papers would have been coded by researchers during LR, the citation relations between these papers are seldom noted. Therefore, the centrality of particular paper(s) in the whole list is largely known. This is the scenario where this recommendation task is supposed to be used. There are two requirements for this task. They are as follows.

Requirement 1 (R1): From the full list of papers collated during the literature review of a particular research project, the researcher would want to identify the important papers based on the centrality in the citation network.

Requirement 2 (R2): The shortlisted papers should vary based on the article-type preference of the user.

4.3.5.2 Citation Network based Shortlisting (CNS) Technique

The technique for this task is called Citation Network based Shortlisting (CNS) technique because the citation network formed with the reading list papers is used to shortlist papers. The technique comprises of two stages. The first stage is the citation clusters stage while the second stage is the paper shortlisting stage. The process flow of the CNS technique is illustrated in Figure 4.7. The two stages are outlined as follows.

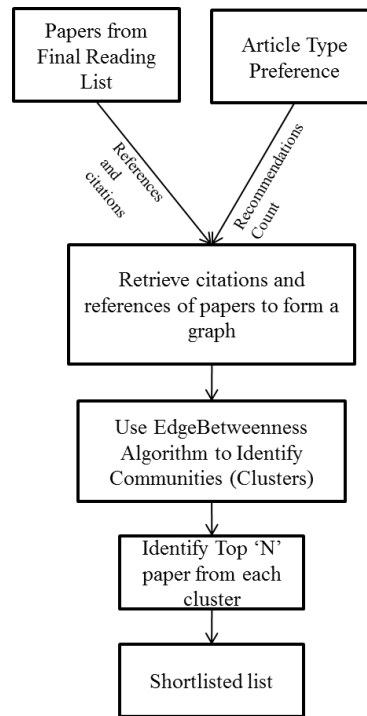


Figure 4.7. Process flow in the CNS technique

Stage 1 – Citation Clusters Formation

The papers in the final reading list are supposedly of the same research area. However, there would be different subsets of papers that cover the various sub-topics in the given research area. The objective of this stage is to identify these clusters/subsets of papers using the citation relations between the papers. For this purpose, the Girvan–Newman algorithm (Girvan & Newman, 2002) is used for identifying the clusters C in the citations network. The algorithm was selected as it is the one of the most prominent community detection algorithms based on link removal. The other algorithms considered were voltage clustering algorithm (Wu & Huberman, 2004) and bi-component DFS clustering algorithm (Tarjan, 1972). Based on internal trail tests, the Girvan–Newman algorithm was able to consistently identify meaningful clusters using the graph constructed with the citations and references of the papers from the reading list.

Stage 2 – Shortlisting Papers from Citation Clusters

The input to this stage is the set of clusters from the community detection algorithm.

In this first version of the CNS technique, the notion of varying the count of shortlisted papers by article-type choice is explored. Therefore, the selection of article-type preference just varies the quantity of shortlisted papers in the final list. For this purpose, four article-types were considered - conference full paper (cfp), conference poster (cp), generic research paper (gp)⁷ and case study (cs). The article-type classification is not part of the ACM metadata but it is partly inspired by the article classification used in Emerald publications⁸. The number of papers to be shortlisted for these article-types was identified by using the historical data from ACM DL dataset which was used as the corpus for the system (details provided in Section 5.2 of Chapter 5). First, the papers in the dataset were filtered by using the title field and section field for the four article-types. Second, the average of the references count was calculated for the filtered papers for each article-type from previous step. The average references count for the article-types gp, cs, cfp and cp were 26, 17, 16 and 6 respectively. This new data is used to set the number of papers to be retrieved from the paper clusters. The final list of shortlisted papers *S* is recommended to the users at the end of the technique. The pseudo-code for the CNS technique is provided below.

1. Initialize the lists *reading_set*, *refcit_set*, *recommendations_set*
2. Input the final reading list papers into the list *reading_set*
3. Input the article-type preference *AT*
4. Retrieve the recommendations count *N* for *AT*
5. Set variable *recommendations_counter* to *N*

⁷ A paper is qualified as a generic research paper when it doesn't fall quality under the requirements of all the other article-types

⁸ Article classification used in Emerald Publications

http://emeraldgrouppublishing.com/products/journals/author_guidelines.htm?id=oir

6. Retrieve the bibliographic references R of the papers in the *reading_set*
7. Retrieve the citations C of the papers in the *reading_set*
8. Populate the list *refcit_set* with R and C
9. Retrieve the citation counts CC of papers in *refcit_set*
10. Create directed graph G with papers from *reading_set* and *refcit_set*
11. Run EdgeBetweenness Algorithm to identify clusters C in graph G
12. Rank the papers in the clusters C based on CC values
13. Set variable *temp_counter* to zero
14. Set variable M to one
15. While *temp_counter* is equal to *recommendations_counter*
 - For each cluster c in C
 - Populate *recommendadation_set* with M ranked paper from the current cluster c
 - Set *temp_counter* to the sum of *temp_counter* and 1
 - Set M to the sum of M and 1 for the current cluster c
 - End For
16. End While
17. Print the list *recommendations_set*

4.4 Rec4LRW System Design

The first objective (RO1) was to identify an appropriate method to map the identified LR and MP tasks to relevant IR/RS algorithms. The assistive system Rec4LRW was developed for addressing RO1. Rec4LRW is an acronym that stands for **Recommender System for Literature Review and Writing**. The singular design principle of the Rec4LRW system is to create an environment meant for effectively supporting the three tasks.

4.4.1 Technical Details

The overall architecture of the Rec4LRW system is illustrated in Figure 4.9. The data loading and curation activities were one-time activities performed when the XML files of the ACM DL extract were loaded into the MySQL base tables. The precomputation of the seven features was also a one-time activity performed on the base tables in order to form aggregated tables. These aggregated tables are the tables on which the SQL queries were executed. The Rec4LRW front-end web pages communicated with corresponding JAVA programs. These JAVA programs in turn executed the SQL queries on the MySQL tables. It is to be noted that all the data displayed in the Rec4LRW system was stored and retrieved from the MySQL database.

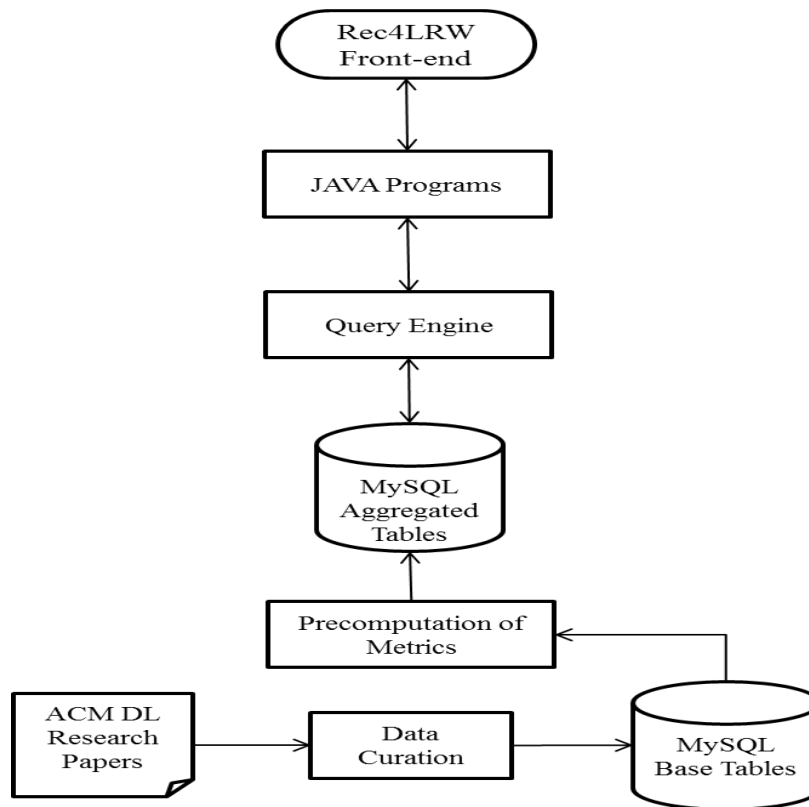


Figure 4.8. Rec4LRW system architecture

In Table 4.3, the libraries and the programming languages used for the Rec4LRW system development are listed. Apache Lucene facilitates content-based

recommendations in Tasks 1 and 2. Apache Mahout was used for implementing CF module in Task 2 while the Jung library was used for executing the Girvan-Newman algorithm in Task 3. Since all the recommendations functionality was coded in Java, PHP was used as the server side scripting interface between the front-end and the recommendation programs. The Rec4LRW server ran on an Intel Xeon E5645 processor with a frequency of 2.40 GHz. The memory (RAM) was 18 GB while the total storage was 1 TB. The server was a windows based system with Windows Server 2008 R2 Enterprise 64-bit as the operation system (OS). The server specification was adequate to handle the traffic and load during the user evaluation study (Study II).

Table 4.3

Technologies and libraries used in Rec4LRW development

Recommendation		
Technique	Web Technologies	Java Library
AKR Technique (Task 1)		Apache Lucene
IDSP Technique (Task 2)	HTML, CSS, JavaScript, PHP	Apache Lucene, Apache Mahout
CNS Technique (Task 3)		Jung

4.4.2 Task Interconnectivity Mechanisms

Task interconnectivity is one of the key aspects of the solution for addressing the second research question. In Rec4LRW, the interconnectivity is enforced through two paper collections – reading list (RL) and seed basket (SB) for the three tasks. Reading list is the list of all papers that are read during literature review. Researcher keeps populating this list as he/she finds new papers relevant to the particular research topic. Reading list is used across all three tasks. Seed basket (SB) is a basket comprising of a

particular set of papers. This set of papers is used to find similar papers as a part of Task 2. SB helps in connecting Task 1 to Task 2 while RL connects all the three tasks in the system.

4.4.3 User-Interface for the Three Tasks

The screenshots of the task screens from the Rec4LRW system are provided from Figures 4.10 to 4.14. The task screens in the first release of the system have been specifically designed for testing purposes. The font aspects (face, color and size) and information positioning of the metadata fields in the task screens were made to resemble the UI of Google Scholar (GS) since GS was the most known system as the findings of Study I. Certain activities are to be performed multiple times by the users since each of the three tasks were evaluated separately. As shown in Figure 4.10 for Task 1, the user has to just select the research topic from the drop-down list box and click on the ‘Generate Recommendations’ button for getting the top 20 paper recommendations.

Rec4LRW - Scientific Paper Recommender System for Literature Review and Writing

Task 1 - Building an Initial reading list of research papers

Please select the research topic:

IMPORTANT NOTE: The papers in the corpus/dataset are from an extract of papers from ACM DL. The below list doesn't include papers indexed in other academic search engines and databases. Please refer the user guide for more information about the dataset used in the system

- Designing a digital library for young children** Survey/Review Popular

Allison Drun, Benjamin B. Bederson, Juan Pablo Hourcade, Lisa Sherman, Glenda Revelle, Michele Platner, Stacy Weng - Digital Libraries, 2001

Abstract: As more information resources become accessible using computers, our digital interfaces to those resources need to be appropriate for all people. However when it comes to digital libraries, the interfaces have typically been designed for older children or adults. Therefore, we have begun to develop a digital library interface developmentally appropriate for young children (ages 5-10 years old). Our prototype system we now call SearchKids offers a graphical interface for querying, browsing and reviewing search results. This paper describes our motivation for the research, the design partnership we established between children and adults, our design process, the technology outcomes of our current work, and the lessons we have learned.

Author Specified Keywords: children; cooperative inquiry; digital libraries; education applications; information retrieval design techniques; intergenerational design team; participatory design; zoomable user interfaces

Citation Count: 23 References Count: 20
- Digital libraries and educational practice: a case for new models** Survey/Review High Reach

Tamara Sumner, Mary Martino - Digital Libraries, 2004

Abstract: Educational digital libraries can benefit from theoretical and methodological approaches that enable lessons learned from design and evaluation projects performed in one particular setting to be applied to other settings within the library network. Three promising advances in design theory are reviewed - reference tasks, design experiments, and design genres. Each approach advocates the creation of 'intermediate' constructs as vehicles for knowledge building and knowledge sharing across design and research projects. One purpose of an intermediate construct is to formulate fine-grained models that describe and explain the relationship between key design features and the cognitive and social dimensions of the context of use. Three models are proposed and used as thought experiments to analyze the utility of these approaches to educational digital library design and evaluation: digital libraries as cognitive tools, component repositories, and knowledge networks.

Author Specified Keywords: cognitive tools; component repositories; design experiments; design genres; design rationale; educational digital libraries; evaluation; knowledge networks; knowledge sharing; reuse

Citation Count: 7 References Count: 52
- Cost and other barriers to public access computing in developing countries** Recent High Reach

Melody Clark, Ricardo Gomez - , 2011

Abstract: Public access to computers and the Internet can play an important role in social and economic development if it effectively helps to meet the needs of underserved populations. Public access venues such as libraries, telecentres and cybercafes are sometimes free, and sometimes charge user fees. User fees can be an important barrier to use of public access venues, especially among underserved communities in developing countries. This paper analyzes the role of user fees and other critical barriers in the use of computers in public access venues in 25 developing countries around the world. Results of this study suggest that digital literacy of staff and local relevance of content may be more important than fees in determining user preference for public access venues. These findings are important to public libraries, which tend to offer free services, but where perceptions of digital literacy of staff and locally relevant content tend to be lowest, compared to telecentres and cybercafes, according to the results of this study. More attention to digital literacy of staff and availability of locally relevant content may be more important than free services to meet the information needs of underserved populations.

Author Specified Keywords: ICT4D; affordability; cybercafes; developing countries; digital literacy; fees; libraries; local content; public access; telecentres

Figure 4.9. Reading list task screen (Task 1)

In the Task 2 screen (refer Figure 4.11), the user has to re-run Task 1 since papers are to be selected for the seed basket. This activity is performed by clicking on the ‘Task 1 papers’ button. A checkbox is provided along with each paper for the user to select the particular paper. After selecting the required number of papers, the user can click on the button ‘Generate Recommendations’. The system picks up the papers from the seed basket as input and generates recommendations as part of Task 2. A sample set of resultant recommendations from Task 2 is show in Figure 4.12. The papers from the seed basket are shown atop the recommendations. In a real-world setting, a researcher would be executing this task multiple times with a different set of input papers until the literature review is complete.

Rec4LRW - Scientific Paper Recommender System for Literature Review and Writing

Task 2 - Finding similar papers based on a set of papers

STEP 1:
Click the below button to regenerate the task1 papers

[Task 1 Papers](#)

STEP 2:
Click the below button to generate recommendations based on the seed basket

[Generate Recommendations](#)

IMPORTANT NOTE: The papers in the corpus/dataset are from an extract of papers from ACM DL. The below list doesn't include papers indexed in other academic search engines and databases. Please refer the user guide for more information about the dataset used in the system

Please add at least 5 papers to the seed basket from the below list

Recommendations for the research topic "digital libraries" (from task 1)

Add/Remove
Paper to
Seed Basket

<input type="checkbox"/>	<p>1) Designing a digital library for young children Survey/Review Popular</p> <p style="font-size: x-small;">Allison Druin, Benjamin B. Bederson, Juan Pablo Hourcade, Lisa Sherman, Glenda Reville, Michele Platner, Stacy Weng - Digital libraries, 2001</p> <p style="font-size: x-small;">Abstract: As more information resources become accessible using computers, our digital interfaces to those resources need to be appropriate for all people. However when it comes to digital libraries, the interfaces have typically been designed for older children or adults. Therefore, we have begun to develop a digital library interface developmentally appropriate for young children (ages 5-10 years old). Our prototype system we now call SearchKids offers a graphical interface for querying, browsing and reviewing search results. This paper describes our motivation for the research, the design partnership we established between children and adults, our design process, the technology outcomes of our current work, and the lessons we have learned.</p> <p style="font-size: x-small;">Author Specified Keywords: children; cooperative inquiry; digital libraries; education applications; information retrieval design techniques; intergenerational design team; participatory design; zoomable user interfaces</p> <p style="font-size: x-small;">Citation Count: 23 References Count: 20</p>		<p>2) Digital libraries and educational practice: a case for new models Survey/Review High Reach</p> <p style="font-size: x-small;">Tamara Sumner, Mary Marlino - Digital libraries, 2004</p> <p style="font-size: x-small;">Abstract: Educational digital libraries can benefit from theoretical and methodological approaches that enable lessons learned from design and evaluation projects performed in one particular setting to be applied to other settings within the library network. Three promising advances in design theory are reviewed - reference tasks, design experiments, and design genres. Each approach advocates the</p>
--------------------------	--	--	---

Figure 4.10. Selecting seed papers before executing Task 2

There are 5 paper(s) in your seed basket

Papers in the seed basket
1) Addressing the challenge of visual information access from digital image and video libraries (2005)
2) Social empowerment and exclusion: A case study on digital libraries (2005)
3) The challenge of virginia banks: an evaluation of named entity analysis in a 19th-century newspaper collection (2006)
4) TableSeer: automatic table metadata extraction and searching in digital libraries (2007)
5) Panorama: extending digital libraries with topical crawlers (2004)

IMPORTANT NOTE: The papers in the corpus/dataset are from an extract of papers from ACM DL. The below list doesn't include papers indexed in other academic search engines and databases. Please refer the user guide for more information about the dataset used in the system

Recommendations based on the above seed basket

- [Faceted metadata for image search and browsing](#) Survey/Review Popular

Ka-Ping Yee, Kirsten Swearingen, Kevin Li, Marti Hearst - Human factors in computing systems, 2003

Abstract: There are currently two dominant interface types for searching and browsing large image collections: keyword-based search, and searching by overall similarity to sample images. We present an alternative based on enabling users to navigate along conceptual dimensions that describe the images. The interface makes use of hierarchical faceted metadata and dynamically-generated query previews. A usability study, in which 32 art history students explored a collection of 35,000 fine arts images, compares this approach to a standard image search interface. Despite the unfamiliarity and power of the interface (attributes that often lead to rejection of new search interfaces), the study results show that 90% of the participants preferred the metadata approach overall, 97% said that it helped them learn more about the collection, 75% found it more flexible, and 72% found it easier to use than a standard baseline system. These results indicate that a category-based approach is a successful way to provide access to image collections.

Author Specified Keywords: faceted metadata; image search interfaces

Citation Count: 101 References Count: 18 [Shared Co-references \(2\)](#) [Shared Co-citations \(1\)](#)
- [How do people manage their digital photographs?](#) Survey/Review Popular

Kerry Rodden, Kenneth R. Wood - Human factors in computing systems, 2003

Abstract: In this paper we present and discuss the findings of a study that investigated how people manage their collections of digital photographs. The six-month, 13-participant study included interviews, questionnaires, and analysis of usage statistics gathered from an instrumented digital photograph management tool called Shoebox. Alongside simple browsing features such as folders, thumbnails and timelines, Shoebox has some advanced multimedia features: content-based image retrieval and speech recognition applied to voice annotations. Our results suggest that participants found their digital photos much easier to manage than their non-digital ones, but that this advantage was almost entirely due to the simple browsing features. The advanced features were not used very often and their perceived utility was low. These results should help to inform the design of improved tools for managing personal digital photographs.

Author Specified Keywords: annotation; content-based image retrieval; digital photography; image browsing; personal photography

Citation Count: 77 References Count: 18 [Shared Co-references \(1\)](#) [Shared Co-citations \(0\)](#)
- [Successful approaches in the TREC video retrieval evaluations](#) Survey/Review Popular

Alexander G. Hauptmann, Michael G. Christel - Multimedia, 2004

Abstract: This paper reviews successful approaches in evaluations of video retrieval over the last three years. The task involves the search and retrieval of shots from MPEG digitized video recordings using a combination of automatic speech, image and video analysis and information retrieval technologies. The search evaluations are grouped into interactive (with a human in the loop) and non-interactive (where the

Figure 4.11. Sample list of recommended papers in Task 2

As mentioned earlier, the third task in the Rec4LRW system is supposed to be executed only during manuscript preparatory stage. At this stage, the researcher knows the type of article that he/she intends to write. The input controls for this task is depicted in Figure 4.13. The article-type preference can be set from the drop-down list box. The available article-types values are conference full paper, conference poster, generic research paper and case study. Along with the article-type, the main input to this task is the reading list papers that are accumulated from the first two tasks. Since the screens have been designed for testing purposes, the user can re-execute Tasks 1 and 2 for adding more papers to the reading list. In Figure 4.14, a sample screenshot of the shortlisted papers is provided.

Task 3 Instructions

Rec4LRW - Scientific Paper Recommender System for Literature Review and Writing

Task 3 - Shortlisting papers from reading list for inclusion in manuscript

OPTIONAL STEP

Click the below button to add more seed papers from task1

[Task 1 Papers](#)

STEP 1:

Click the below button to re-run task 2 for adding papers to your reading list

[Task 2 Papers](#)

STEP 2:

Select the article-type of your manuscript

conference full paper ▼

STEP 3:

Click the below button to generate recommendations based on the reading list

[Generate Recommendations](#)

There are only 19 papers in your reading list. Please make sure the list has atleast 30 papers. You can either re-run task 2 by clicking the button 'Task 2 papers' in this page for adding more papers or you can select a different set of seed papers by clicking on 'Task 1 papers'

Figure 4.12. Input options in Task 3

Shortlisted papers based on the article-type preference

1) [SIA: secure information aggregation in sensor networks](#) Popular

Bartosz Przydatek, Dawn Song, Adrian Perrig - Embedded networked sensor systems, 2003

Abstract: Sensor networks promise viable solutions to many monitoring problems. However, the practical deployment of sensor networks faces many challenges imposed by real-world demands.

Author Specified Keywords: approximate information aggregation; interactive proofs; security; sensor networks

Citation Count: 44 References Count: 27 [View papers in the parent cluster](#)

2) [The UCON_{Ag} usage control model](#) Popular

Jaeheung Park, Ravi Sandhu - ACM Trans. Inf. Syst. Secur., 2004

Abstract: In this paper, we introduce the family of UCON_{Ag} models for usage control (UCON), which integrate Authorizations (A), oBligations (B), and Conditions (C). The term usage control is a generalization of access control to cover authorizations, obligations, conditions, continuity (ongoing controls), and mutability.

Author Specified Keywords: access control; digital rights management; privacy; trust; usage control

Citation Count: 31 References Count: 43 [View papers in the parent cluster](#)

3) [Role-based access control for publish/subscribe middleware architectures](#) Popular

András Belkissoszló, David M. Eysers, Peter R. Pietzuch, Jean Bacon, Ken Moody - Distributed event-based systems, 2003

Abstract: Research into publish/subscribe messaging has so far done little to propose architectures for the support of access control, yet this will be an increasingly critical requirement as systems move to Internet-scale. This paper discusses the general requirements of publish/subscribe systems with access control. We then present our specific integration of OASIS role-based access control into the Hermes publish/subscribe middleware platform. Our system supports many advanced features, such as the ability to work within a network where nodes are attributed different levels of trust, and employs a variety of access restriction methods which balance expressiveness with the content-based routing optimisations available. We illustrate our achievements by discussing an application scenario in which our system will be of particular use.

Author Specified Keywords: broker trust; publish/subscribe; restriction of advertisements/subscriptions; role-based access control

Citation Count: 45 References Count: 10 [View papers in the parent cluster](#)

4) [PSFQ: a reliable transport protocol for wireless sensor networks](#) Popular

Chieh-Yih Wan, Andrew T. Campbell, Lakshman Krishnamurthy - Wireless sensor networks and applications, 2002

Abstract: We propose PSFQ (Pump Slowly, Fetch Quickly), a reliable transport protocol suitable for a new class of reliable data applications emerging in wireless sensor networks. Due to the application-specific nature of sensor networks, it is difficult to design a single monolithic transport system that can be optimized for every application.

Author Specified Keywords: reliable transport protocols; wireless sensor networks

Citation Count: 39 References Count: 14 [View papers in the parent cluster](#)

5) [ESRT: event-to-sink reliable transport in wireless sensor networks](#) Popular

Yogesh Sankarasubramanian, Ozgur B. Akan, Ian F. Akyildiz - Mobile ad hoc networking & computing, 2003

Abstract: Hence, conventional end-to-end reliability definitions and solutions are inapplicable in the WSN regime and would only lead to a waste of scarce sensor resources. To the best of our knowledge, reliable transport in WSN has not been studied from this perspective before. In order to address this need, a new reliable transport scheme for WSN, the event-to-sink reliable transport (ESRT) protocol, is presented in this paper.

Author Specified Keywords: congestion control; energy conservation; event-to-sink reliability; reliable transport protocols; wireless sensor networks

Citation Count: 31 References Count: 13 [View papers in the parent cluster](#)

6) [Sensor networks for medical care](#)

Victor Shnayder, Bor-rong Chen, Konrad Lorincz, Thaddeus R. F. Fulford Jones, Matt Welsh - Embedded networked sensor systems, 2005

Abstract: No data

Author Specified Keywords: medical sensor networks; sensor query processing; wireless routing; wireless sensor networks

Citation Count: 28 References Count: 3 [View papers in the parent cluster](#)

Information Cue Labels

Parent cluster of the shortlisted paper

Time taken for shortlisting articles = 6 seconds

[Click here to start evaluation of this task](#)

Figure 4.13. Sample list of shortlisted papers in Task 3

4.4.4 Informational Display Features in the Rec4LRW User-Interface

The information displayed as part of the recommendations is as critical as the quality of the papers as the relevance judgment on the relevance of the papers is made based on the displayed metadata. In previous SPRS studies (Ekstrand et al., 2010; Jardine, 2014; Mcnee, 2006) where user evaluation exercises have been performed, only the basic metadata information (paper citation, year, authors, title, publication name/type) have been displayed. In Table 4.4, the metadata fields displayed as part of the informational display features in the Rec4LRW system are listed. A column to indicate whether the feature is a traditional or novel feature has been included in the table.

Table 4.4

Informational display features in Rec4LRW task screens

Display Feature	Traditional/New	Tasks
Title, Authors, Publication Year, Summary/Abstract,	Traditional	All

Display Feature	Traditional/New	Tasks
Author-Specified Keywords, Citations Count, References Count		Tasks
Information Cue Labels	New	All Tasks
Word Cloud of Author-Specified Keywords	New	Task 1
Shared Co-Citations and Co-References with Seed Basket Papers	New	Task 2
View Papers in the Parent Cluster	New	Task 3

4.4.3.1 Novel Display Features

The display philosophy in the Rec4LRW system was to include features that accentuate the characteristics of the recommendation techniques used for the three tasks. Since there are redesigned aspects in the tasks, certain new informational features were conceptualized and included along with the other basic metadata fields in the task screens. They are described as follows.

Information Cue Labels

In Task 1, there are four types of papers (recent, diverse, survey and popular papers) retrieved for the input research topic. If these types are displayed as an informational feature, it would help the user in differentiating the papers in the recommendations list. Information cues offer the potential to impact the user perception of retrieved results, an observation seen in past studies (Tang, 2009). The usage of cue labels is new to academic search systems although its effectiveness has been proved in other domains (Verbeke & Ward, 2006).

For the Rec4LRW system, four labels were included aside the paper title in the user-interface. The display logic for the cue labels are described as follows. The *recent* label is displayed for papers published between the years 2009 and 2011 (the most recent papers in the ACM dataset is of 2011). The *survey/review* label is displayed for papers which are of the type - literature survey or review. For the *popular* label, the unique citation counts of all papers for the selected research topic are first retrieved from the database. The label is displayed for a paper if the citation count is in the top 5% percentile of the citation counts for that topic. Similar logic is used for the *high reach* label with references count data. These labels are displayed in an automated mechanism based on the paper metadata. The labels and display intent are listed in Table 4.5. The cue labels alongside the recommended paper titles are shown in the screenshot in Figure 4.15.

Table 4.5

Information cue labels and their display intents

Label	Display Intent
Recent	Indicate recently published papers
Popular	Indicate the paper is highly cited for the input research topic
High Reach	Indicate the paper has a high number of references in its bibliography
Survey/Review	Indicate literature survey or review papers



Figure 4.14. Information cue labels beside the paper titles in the task recommendations

Word Cloud of Author-specified Keywords

The diversity characteristic in Task 1 is meant to indicate that the reading list comprises of papers covering a wide variety of sub-topics of the input research area. If the sub-topic is displayed as cue label along with the paper title, it might not be useful for researchers who are new to the particular research area. Instead, a word cloud generated with the author-specified keywords of the retrieved papers would be more beneficial for two reasons. First, it provides a snapshot of the sub-topics covered by the researcher papers. Second, it differentiates the more frequent sub-topics from the less frequent topics with the use of font sizes (Verbeke & Ward, 2006). A sample word cloud generated for the 'Information Retrieval' is depicted in Figure 4.16. The word cloud feature is displayed in the bottom of the screen, right after the recommendations list. It is to be noted that this word cloud feature is an alternative representation of the author-specified-keywords of the retrieved papers from Task 1. Unlike the original textual display, this feature helps the user in gaining faster insights about the retrieved papers.

1) [Design space exploration using time and resource duality with the ant colony optimization](#) Popular
 Gang Wang; Wenrui Gong; Brian DeRenzi; Ryan Kastner - , 2006
Abstract: Design space exploration during high level synthesis is often conducted through ad-hoc probing of the solution space using some so dependent on designer's experience. We propose a novel design exploration method that exploits the duality between the time and resource of a high quality time/area tradeoff curve in a fast, effective manner. It uses the MAX-MIN ant colony optimization to solve both the time and resource constrained algorithms to quickly traverse the design space. Compared to using force directed scheduling exhaustively at every time (average 17.3% reduction of resource counts) with similar run time on a comprehensive benchmark suite constructed with classic and real-life problem sizes.
Author Specified Keywords: MAX-MIN ant system; ant colony optimization; design space exploration; instruction scheduling algorithms
 Citation Count: [69](#) References Count: [14](#) [Shared Co-references \(0\)](#) [Shared Co-citations \(0\)](#)

2) [A threshold of \$\ln n\$ for approximating set cover](#) Popular
 Uriel Feige - J. ACM, 1998
Abstract: Given a collection F of subsets of $S = \{1, \dots, n\}$, setcover is the problem of selecting as few as possible subsets from F such that their union has maximum cardinality. Both these problems are NP-hard. We prove that $(1 - o(1)) \ln n$ is a threshold below which setcover slightly superpolynomial time algorithms. This closes the gap (up to low-order terms) between the ratio of approximation achievable by the greedy algorithm, that showed hardness of approximation within a ratio of $\log 2n/2 \approx 0.72 \ln n$. For maxk-cover, we show an approximation threshold of $(1 - o(1)) \ln n$.
Author Specified Keywords: approximation ratio; set cover
 Citation Count: [57](#) References Count: [34](#) [Shared Co-references \(8\)](#) [Shared Co-citations \(2\)](#)

Figure 4.16. Share co-references and co-citations features in Task 2 recommendations

Clustered Papers

In Task 3, the objective is to identify unique and popular papers from the researcher's entire reading list. As described earlier, community detection algorithm was used for identifying clusters of related papers. Subsequently, the top cited papers are recommended from each cluster. In such a scenario, the researcher would be keen to know the parent clusters from which the recommended paper is shortlisted. Therefore, a corresponding informational feature is required for this purpose. The display can be both textual or in the form of a network. In the current research, the information is displayed in a textual form. The 'View papers in the parent cluster' feature alongside the references and citations count is shown in the screenshot in Figure 4.18.

1) [Information retrieval on the semantic web](#) Popular
 Urvi Shah; Tim Finin; Anupam Joshi; R. Scott Cost; James Matfield - Information and knowledge management, 2002
Abstract: We describe an approach to retrieval of documents that contain of both free text and semantically enriched both documents and queries can be marked up with statements in the DAML+OIL semantic web language. These statements are indexed, when a query is processed and when query results are evaluated.
Author Specified Keywords: hybrid information retrieval; query-answering systems; semantic web; text extraction
 Citation Count: [223](#) References Count: [28](#) [View papers in the parent cluster](#)

2) [Information retrieval on the web](#) Survey/Review Popular High Reach
 Mei Kobayashi; Koichi Takeda - ACM Comput. Surv., 2000
Abstract: In this paper we review studies of the growth of the Internet and technologies that are useful for information e.g., current as well as projected number of users, hosts, and Web sites. Although numerical figures vary, overall trend is that the growth is rapid. Hence it is not surprising that about 85% of Internet users surveyed claim using search engines and search engines with the performance of the current generation of search engines; the slow retrieval speed, communication delays, and discuss the development of new techniques targeted to resolve some of the problems associated with Web-based information retrieval.
Author Specified Keywords: Internet; World Wide Web; clustering; indexing; information retrieval; knowledge management
 Citation Count: [36](#) References Count: [231](#) [View papers in the parent cluster](#)

Figure 4.17. Parent cluster papers viewing feature in Task 3 recommendations

4.5 Summary

The base features and the task interconnectivity mechanisms in the Rec4LRW system constitute the solution to the second research question. The system design of the Rec4LRW system highlights the minimalist design principle with a high level of resemblance to Google Scholar. The novel informational display features of the system were meant to highlight the redesigned task aspects for the three tasks. The task redesign, novel informational display features and the task interconnectivity mechanisms are collectively known as the threefold intervention framework. This framework can be used in future SPRS studies since it relates the redesigned task features to the user-interface elements in an environment where multiple tasks are to be supported. RQ2 and RO1 were addressed in this chapter through conceptualization of the recommendation techniques of the three tasks and design of the assistive system respectively. During the design phase of the Rec4LRW system, few optimization tweaks were performed to improve the algorithmic performance and the overall system reaction time (details provided in Appendix C).

CHAPTER FIVE - STUDY II: EVALUATION OF RECOMMENDATION TECHNIQUES AND REC4LRW SYSTEM

5.1 Introduction

The evaluation studies conducted as part of Study II in Phase 3 are presented in this chapter. First, the offline evaluation experiment conducted for the AKR technique is presented. Details regarding the user evaluation study conducted with the Rec4LRW system are next presented. The data analysis of the user evaluation study is presented in two segments. The first segment reports the quantitative analysis of the data collected through survey-type questions. In the second segment, observations from the analysis of the participants' feedback data (qualitative data) are presented. This chapter ends with a conclusion section in which the major findings from the evaluation studies are summarized.

Study II was designed to address the research objective RO2 of evaluating the developed system. Therefore, the overall objective of Study II was to evaluate the recommendation techniques of the three tasks from user and system perspective. A pre-study was conducted in the form of an offline evaluation experiment for Task 1's AKR technique since this task has been partially addressed in previous SPRS studies. The main part of Study II was the user evaluation study where users evaluated the three tasks along with overall system for different evaluation metrics.

5.2 Dataset used for Study II

The proposed recommendation techniques for the three tasks required the following fields of a research paper – title, abstract, publication year, author(s), publication venue, author-specified keywords, categories, bibliographic references and citations.

Therefore, information sources selection was constrained by the availability of these fields in the datasets. Out of the surveyed information sources, only ACM DL (ACM Digital Library) dataset provided all the required fields. The field ‘author-specified keywords’ was conspicuously missing in all the other datasets. Therefore, the decision was made to use the ACM DL dataset for the Rec4LRW system. A research agreement was signed with ACM in August 2013 for establishing the rules and regulations on the dataset usage. ACM provided the data in the form of XML files. A total of 12,177 files were sent via FTP from the ACM server. The XML files were grouped in two sections – periodicals and proceedings. Periodicals refer to papers published in journal, magazines and transactions of ACM DL while proceedings refer to papers published as part of conferences, symposiums and workshops. Since a MySQL database was used as the back-end database for the Rec4LRW system, the data from the XML files were parsed, partially transformed and loaded in relational tables in the database using JAVA programs. Empty XML files and redundant files were discarded during this load process. Papers in the dataset were published for the period 1951 to 2011.

The papers for the Rec4LRW system were shortlisted based on full text and metadata availability in the provided dataset as there were papers with missing data. The final sample set contained a total of 103,739 articles and corresponding 2,320,345 references. Even though, the references were provided as separate XML tags for each paper in the dataset, the text had to be further parsed to identify the reference paper title, publication year and venue. The references were parsed using AnyStyle parser (AnyStyle, 2015).

5.3 Pre-study: Offline Evaluation of AKR Technique

5.3.1 Evaluated Techniques

Task 1's AKR technique (refer Section 4.3.3.2 of Chapter 4) is a unique recommendation technique as it considers four different paper-types for generating the reading list. It has been highlighted in Section 2.3.3.1 of Chapter 2 that item-based collaborative filtering (IBCF) boosted with graph ranking algorithms such as PageRank (Page et al., 1999) and HITS (Kleinberg, 1999) have provided better results than content-based RS filtering techniques (Ekstrand et al., 2010). Both these approaches were considered for the present experiment. The third approach considered was the traditional PageRank technique.

The proposed AKR technique was evaluated through four variants. The composite rank in Stage 2 of the AKR technique is based on three metrics – *TPC*, *citation count* and *reference count*. Hence, different weights can be set to these metrics. In the first set of variants, the basic AKR technique was implemented with two different weights combinations. The first variant AKRv1 had the weights set ($W_{CC} = 0.25$, $W_{RC}=0.25$, $W_{CO} = 0.5$) where *TPC* is given 50% weightage while *citation count* and *reference count* share the remaining 50% weightage. This weight set provides equal weightage to both relative importance (*TPC*) and global importance (*citation count*, *reference count*) of a given paper in the citation network. Second variant AKRv2 had the weights set ($W_{CC} = 0.1$, $W_{RC}=0.1$, $W_{CO} = 0.8$) where *TPC* is given a dominant 80% importance while *citation count* and *reference count* share the remaining 20% weightage. This weight set provides higher weightage to relative importance (*TPC*) of a given paper in the citation network. The weights are set in a slightly arbitrary fashion for this second variant. For instance, the weights set ($W_{CC} =$

0.1, $W_{RC}=0.1$, $W_{CO} = 0.8$) can also be ($W_{CC} = 0.15$, $W_{RC}=0.15$, $W_{CO} = 0.7$). Anyhow, these weight set combinations highlight the flexibility in the AKR technique. There are two other alternative options for setting the weight values. The first option is conducting a separate experiment with a particular threshold while the second option is allowing the users to directly control these weights. The latter option can be considered as the most ideal option if the users get to control these weights using range slider widgets in the user-interface of the system. The caveat is that users are required to learn the usability of these weights and their impact on the final output. Also, users with limited experience, may instead expect the system to generate the optimal output instead of having to set the weight values themselves.

The second set of AKR variants were introduced after considering the findings from a case study which was performed with a single research topic. The case study is presented in Appendix E. The case study was conducted primarily to observe AKR technique's performance. It was observed that *TPC*'s effectiveness in meeting the requirements of a reading list would be enhanced if it is boosted with HITS score. Therefore, the third and fourth variants of AKR technique had the final ranking score calculated by combining the composite rank and HITS value of a given paper. The third AKR variant called the HAKRv1 technique (HITS enhanced AKR) had the same weights set as AKRv1 and the fourth variant HAKRv2 had the same weights set as AKRv2. The seven techniques used for the offline evaluation are listed in Table 5.1.

Table 5.1

Techniques used in offline evaluation experiment

Order	Abbr.	Technique Description
A	AKRv1	Basic AKR technique with weights $W_{CC} = 0.25$, $W_{RC}=0.25$, $W_{CO} =$

Order	Abbr.	Technique Description
		0.5
B	AKRv2	Basic AKR technique with weights $W_{CC} = 0.1$, $W_{RC}=0.1$, $W_{CO} = 0.8$
C	HAKRv1	HITS enhanced AKR technique boosted with weights $W_{CC} = 0.25$, $W_{RC}=0.25$, $W_{CO} = 0.5$
D	HAKRv2	HITS enhanced AKR technique boosted with weights $W_{CC} = 0.1$, $W_{RC}=0.1$, $W_{CO} = 0.8$
E	CFHITS	IBCF technique boosted with HITS
F	CFPR	IBCF technique boosted with PageRank
G	PR	PageRank technique

5.3.2 Experiment Setup

The top 200 author-specified keywords from the ACM DL dataset were identified as the seed research topics for the experiment. This selection of the keywords was based on the count of papers where particular keyword was one of the five keywords specified by the authors in the metadata. 14 keywords were removed as these keywords were duplicates. A final total of 186 keywords were used for the experiment. The list of research topics is provided in Appendix G.

The experiment was performed in three sequential steps. The first and third steps were made common for all the seven techniques. In the first step, top 200 papers were retrieved using the BM25 similarity algorithm. In the second step, the top 20 papers were identified using the specific ranking schemes of the seven techniques. In the final step, the evaluation metrics were measured for the seven techniques.

5.3.3 Evaluation Metrics

Traditional evaluation metrics such as Precision, Recall and DCG (Discounted Cumulative Gain) were not used for the evaluation as the intent in this experiment was to identify the approach which best satisfies the four requirements.

For the reading list requirements R1, R2 and R3 (refer Section 4.3.3.1 of Chapter 4), the number of relevant papers were enumerated as a part of the metrics measurement. For requirement R4 (diverse papers), a different approach was required. In earlier RS studies, measurement of novelty and diversity metrics has been based on distance between the recommended items (Castells, Vargas, & Wang, 2011). Longer distance between two recommended items indicated high diversity while shorter distance indicated low diversity. In this experiment's context, since there are references and citations networks for the papers, subgraph properties could be used to infer the level of diversity. If $G(V,E)$ is a graph built with references and citations of papers for a particular topic, $GI(VI,EI)$ is a subgraph built with just the final 20 papers from the evaluated retrieval technique. The number of edges EI from GI is an indication of level of diversity in the final list of papers. If there is more number of edges, it means there are many inter-referencing/inter-citing connections between the papers, thus implying a less diverse list of papers and vice-verse for high diversity. Hence, the number of edges was the main criterion for R4.

The evaluation rhetoric in this offline evaluation experiment is explained as follows. For each of the four requirements (except R4), the number of X papers were first identified in the final top 20 list (for instance, X being recent, popular or survey paper). The paper count was compared across the seven techniques. Ranks were assigned to the technique based on the highest counts. This step was repeated for all

the 186 topics. There were ranked lists for 186 topics for each requirement. Rank aggregation approaches (Dwork, Kumar, Naor, & Sivakumar, 2001) have been proposed to identify the optimal ranked list of items over multiple lists. In this experiment, the RankAggreg library (Pihur & Datta, 2009) was employed to identify the optimal ranked lists for the four requirements. Specifically, Cross-Entropy Monte-Carlo algorithm (CE) with two distance measures Spearman footrule distance and Kendall's tau distance was used for the rank aggregation. Based on the aggregated ranks for the four requirements, the best technique which satisfies the most requirements was finally selected.

5.3.4 Findings

The ordering for the seven techniques is provided in Table 5.1. The optimal aggregated ranked lists and minimum objective scores for the two distance measures are displayed in Tables 5.2 and 5.3. The lowest score (minimum objective function score) was achieved by the CE algorithm run with Kendall distance. The comparison between spearman distance, mean ranks and samples of raw data is provided in Figures 5.1-5.4. The figures help in comparing the optimal rank lists with the mean (average) ranking method, for each paper type. In the X axis of the figures, the techniques are listed in the optimal order so that the distinct ranks can be observed. In the Y axis of the figures, the ranks are displayed. From the figures, the drawback of using a simple averaging method in rank aggregation can be noticed since certain techniques share the same ranks.

Table 5.2

Aggregated ranks generated using CE algorithm with spearman footrule distance

Paper Type (Requirement)	Optimal Aggregated Ranks							Min. Obj. Function Score
	1	2	3	4	5	6	7	
Recent Papers (R1)	B	A	C	D	E	F	G	10.66
Popular Papers (R2)	F	E	C	D	G	A	B	11.89
Literature Survey Papers (R3)	C	G	D	A	E	F	B	13.38
Diverse Papers (R4)	C	D	G	A	B	F	E	12.15

Note: The optimal ranked list is displayed for the paper type in each row. For instance, the AKR variants occupy all the top ranks for the requirement R1. The letters corresponding to the seven techniques are listed in Table 5.1.

Table 5.3

Aggregated ranks generated using CE algorithm with kendall distance

Paper Type (Requirement)	Optimal Aggregated Ranks							Min. obj. Function Score
	1	2	3	4	5	6	7	
Recent Papers (R1)	A	B	C	D	E	F	G	6.30
Popular Papers (R2)	F	C	E	D	G	A	B	7.31
Literature Survey Papers (R3)	C	D	G	E	A	F	B	8.10
Diverse Papers (R4)	C	D	G	A	E	B	F	7.37

Note: The optimal ranked list is displayed for the paper type in each row. For instance, the AKR variants occupy all the top ranks for the requirement R1. The letters corresponding to the seven techniques are listed in Table 5.1.

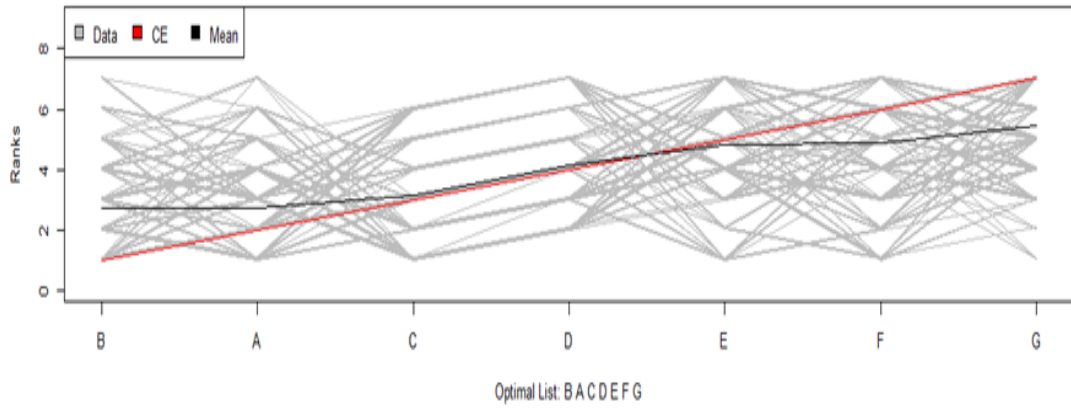


Figure 5.1. Recent papers rank aggregation comparison

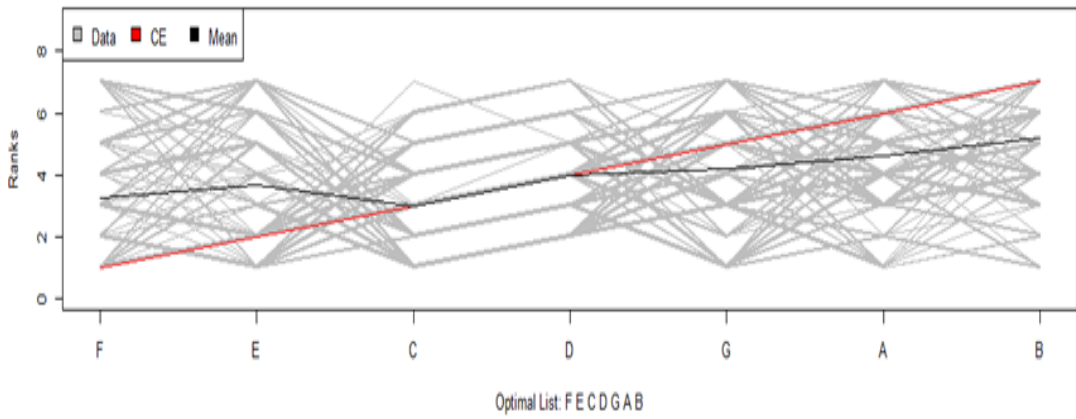


Figure 5.2. Popular papers rank aggregation comparison

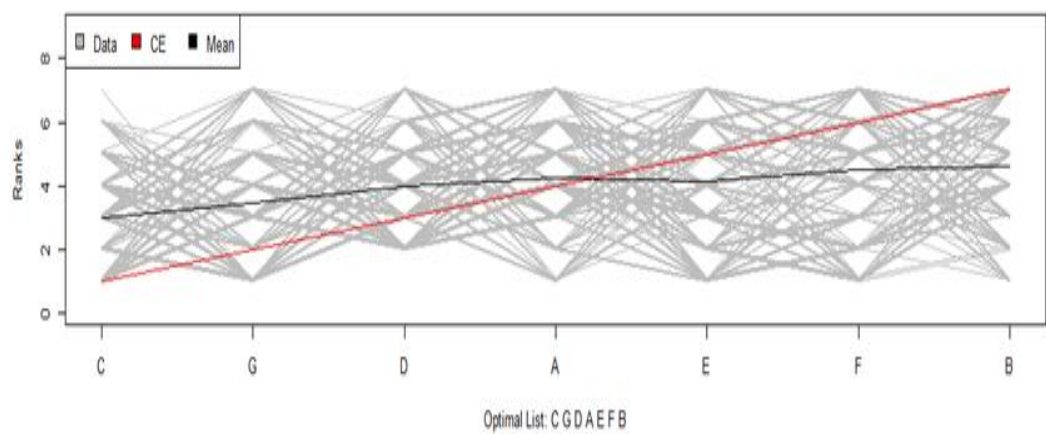


Figure 5.3. Literature survey papers rank aggregation comparison

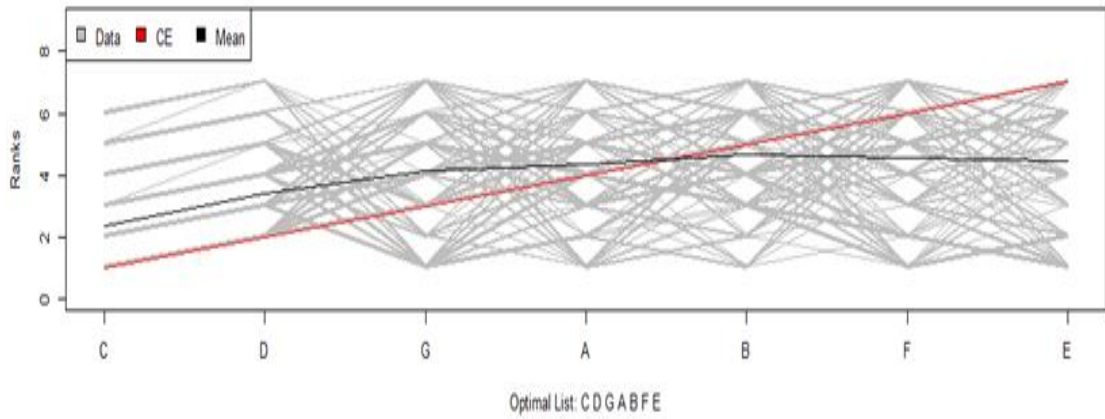


Figure 5.4. Diverse papers rank aggregation comparison

Both the distance measures (Spearman and Kendall) had similar results in the context of top three ranks. Table 5.3 is used for the interpretation of the final results. For R1, the basic variants of the AKR technique (A & B) provided the best results with most number of recent papers. The HITS enhanced AKR variants (C & D) were the next best entries in the ranked list. The basic AKR techniques consider both references and citations count by design. Therefore, recent papers with high number of references have an increased probability for being shortlisted when compared to other techniques. The benchmarked techniques (E, F & G) are not designed to retrieve recent papers and hence, the low ranks for these techniques.

For R2, the benchmarked techniques, particularly CFPR (F) provided the best results, thereby validating their usefulness in finding seminal/classical papers, similar to previous study results (Ekstrand et al., 2010). Among the AKR variants, the HAKRv1 technique (C) was within the top 3 ranks mainly because of the influence of paper's HITS score on the *TPC* value, thereby validating the main purpose behind boosting *TPC* with HITS score. There was a perceivable gap between HAKRv1 (C) and HAKRv2 (D) in this case. The former variant where 50% weightage is given to

TPC as against 80% in the latter variant, clearly benefits papers with higher citation counts.

For R3, HAKRv1 technique (C) yields the best result because of the ability of *TPC* in giving precedence to papers with bigger bibliographies. Interestingly, the basic AKR techniques rank lower in the list, which may be due to lack of recent literature survey papers in the collection. For R4, HAKRv1 technique (C) gets the top rank again due to lesser degree of connections between the papers in the final list. All the variants of AKR are supposed to have lesser connections as the aim is to produce a diverse list of papers. However, the PR technique (G) seems to perform better than basic AKR variants (A & B).

From the findings, it is evident that HAKRv1 (C) is the top performing technique among the seven approaches, since it best satisfies the four requirements. This technique acquired the top most rank for R3 and R4 while it was within the top 3 ranks for R1 and R2. It is to be noted that HAKR variants (C&D) closely follow each other in three requirements with R2 as the only exception. The weight set combinations were important in distinguishing the performance of the variants. For all the four AKR variants, the first weight set ($W_{CC} = 0.25$, $W_{RC} = 0.25$, $W_{CO} = 0.5$) provided better results than the second weight set ($W_{CC} = 0.1$, $W_{RC} = 0.1$, $W_{CO} = 0.8$). This observation validates the importance of two absolute values of a research paper – references and citations count. *TPC* is a relative value since it is based on the author-specified keywords. The intuition to include all these three values in the composite rank was to consider both absolute and relative values of a research paper in the final ranking step. If *TPC* had been directly used for ranking the papers in the final step, the requirements R2 and R3 would have been affected as lesser number of papers would have been shortlisted in the top 20 papers. With this experiment, the extensibility of

the *TPC* in AKR techniques is shown. HITS score was used to enhance the ability of the AKR technique in obtaining top ranks for all the four requirements. As a result, the HAKRv1 technique was the sole technique selected for the user evaluation study. This pre-study helped in addressing RQ3 of the current research through the completion of the offline evaluation.

5.4 User Evaluation Study

In this section, the user evaluation study conducted as part of Study II is described. After the presentation of the objectives and participant demographics, the evaluation results are presented for each of the three tasks followed by the system-level results. Under each sub-section, the evaluation measures are first introduced followed by the findings.

5.4.1 Objectives

Research on RS typically focuses on the accuracy of the algorithms. However, accuracy only partially constitutes the user experience of a RS. The premise of algorithm based offline evaluation is that better algorithms lead to better quality recommendations, which could lead to better user experience. Nevertheless, many studies (Konstan & Riedl, 2012; McNee, Riedl, & Konstan, 2006) have argued that other factors influence the user experience and that these factors have not received much attention in SPRS evaluation (Joeran Beel, Genzmehr, et al., 2013). Therefore, the importance of user-based evaluation is crucial towards acceptance of RS, regardless of the domain.

The purpose of the user evaluation study in Study II was to determine whether researchers using the tasks provided by Rec4LRW system can be efficient and

effective in conducting the corresponding LR tasks. Researchers' perceptions of the individual characteristics of the recommended papers, overall quality of the recommendation list and system features were measured. For comparison purposes during analysis, the participants were split into two groups – students and staff. Students group comprise of participants who have identified themselves as graduate research students. Staff group comprise of participants who have identified themselves as a teaching faculty, research staff or academic/research librarians. It has been already highlighted from information behavior studies (Section 2.2.3.1 in Chapter 2) that graduate research students suffer the most during information seeking of research papers. Hence, there is a necessity for grouping an appropriate set of participants as a part of the students group. On the other hand, the staff group can also be considered as a non-students group in the context of this research. It is to be noted that there is no logical concern in combining non-PhD staff (research assistants, research associates, research engineers and librarians) along with teaching faculty in the staff group. This statement has been made on the lack of scientific information behavior studies that have reported the issues faced by non-PhD research staff. Hence, there is no claim that these types of users are also facing issues with literature review or scientific information seeking in general. Therefore, the analysis of the evaluation data would be best served by splitting the participants into the two aforementioned groups (students and staff). The specific evaluation goals for this study were as follows.

- i. Ascertain the agreement percentages of the evaluation measures for the three tasks and the overall system and identify whether the values are above a preset threshold criteria of 75%. The agreement percentage based evaluation approach has been adopted from (Jardine, 2014).

- ii. Test the hypothesis that students benefit more from the recommendation tasks/system in comparison to staff. It is to be noted that hypothesis testing (Snedecor & Cochran, 1989) among different groups of participants has not been performed in earlier SPRS studies.
- iii. Measure the correlation between the measures and build a regression model with ‘agreeability on a good list’ as the dependent variable. Similar to the previous evaluation goal, correlation analysis (Myers, Well, & Lorch, 2010) and multiple regression modeling (Cohen, Cohen, West, & Aiken, 2003) has not been attempted in prior SPRS studies
- iv. Track the change in user perceptions between the three tasks. This is similar to the first evaluation goal since the agreement percentages will be used for the analysis.
- v. Compare the pre-study and post-study variables for understanding whether the target participants are benefitted from the tasks. This goal is executed using a crosstab analysis (Gokhale & Kullback, 1978) between the pre-study and post-study variables.
- vi. Identify the top most preferred and critical aspects of the task recommendations and the system using the subjective feedback of the participants.

5.4.2 Participant Recruitment and Screening

A directed participant recruitment approach was employed as the participants were required to have research paper authoring experience. Three communication channels were used for advertising the study. Invitation mails were sent to all postgraduate students and staff of three schools in Nanyang Technological University. The three schools are school of electrical and electronic engineering (EEE), school of computer

science and engineering (SCSE) and school of communication & information (SCI). These schools were selected since the papers in the ACM DL dataset were applicable mainly to the research conducted in these schools. Similar mails were sent to a selected set of postgraduate students and staff of electrical, electronics, computer science and information systems schools in National University of Singapore (NUS), Singapore Management University (SMU) and Singapore University of Technology and Design (SUTD). This selected set was collated by scraping data from the websites of these three universities. Secondly, advertisement posters were put up in notice boards across Nanyang Technological University. Finally, invitation mails were also sent to mailing lists related to Library and Information Studies (LIS) and Information Systems.

The main selection criteria was that participant should have authored at least one conference or journal paper. A pre-screening survey was conducted to shortlist the potential participants. The whole study was conducted from second week of November 2015 to end of January 2016. The Rec4LRW system was made available through the internet so that the user evaluation study could be conducted. Participants were permitted to perform the study from any location. A total of 230 researchers completed the pre-screening survey.

5.4.3 Participant Demographics

230 researchers expressed interest for participating in the study, out of which 211 participants satisfied the entry criteria of the study. Out of the eligible 211 participants, 149 participants signed the consent form. 119 of them completed the whole study inclusive of the three tasks in the system. Since, the guidelines allowed the participants to leave the study at any point, participants left the study at different points of time. In

Table 5.4, the number of participants who completed the evaluation at different stages is listed.

Table 5.4

Number of participants by stage

Stage	No. of users who completed the evaluation
Task 1 - Building an initial reading list of research papers	132
Task 2 - Finding similar papers based on set of papers	121
Task 3 - Shortlisting articles from reading list for inclusion in manuscript	119

The participant demographics breakdown is based on the 132 participants who completed at least the first stage of evaluation i.e. Task 1 evaluation. The number of participants and the corresponding percentages are provided in Table 5.5 for the different demographic variables. 62 participants were PhD/MSc students while 70 were research staff, academic staff and librarians. Among the staff, participants with the position research fellow (13.6%), research associate (8.3%) and assistant professor (7.6%) had the highest representation in the overall sample. The average research experience for PhD students was 2 years while for staff, it was 5.6 years. This sizeable difference in the research experience helps in segregating the two user groups in an adequate manner and it is also a good indicator for the validity of the subsequent results. However, majority of the participants claimed they had intermediate experience level (46.2%) and only a few participants claimed they were beginners

(11.4%). Therefore, this perceived research experience variable could not be alternatively used for categorizing the participants into two different groups.

Majority of the participants were from the engineering disciplines (65.9%) with nearly 39% of the overall participants specifically from the computer science discipline. Library and information studies (LIS) and electrical disciplines were also well represented with 30 participants (22.7%) each. Since the study was well-advertised within the university, 87 (65.9%) participants were from Nanyang Technological University (NTU) and 45 (34.1%) were from other universities. This sample turned out to be an exclusive Asian sample since 124 (93.9%) participants were from Asian countries and only 8 (6%) participants were from countries outside of Asia.

Table 5.5

Participant demographic variables

Demographic Variable		Number of Participants
<i>User Group</i>		
	Student	62 (47%)
	Staff	70 (53%)
<i>Position</i>		
	Graduate Research Student	62 (47%)
	Research Fellow	18 (13.6%)
	Research Associate	11 (8.3%)
	Assistant Professor	10 (7.6%)
	Librarian	9 (6.8%)
	Associate Professor	6 (4.5%)

Demographic Variable	Number of Participants
Lecturer	4 (3%)
Research Engineer	3 (2.3%)
Senior Research Scientist	2 (1.5%)
Senior Lecturer	2 (1.5%)
Research Assistant	2 (1.5%)
Project Officer	2 (1.5%)
Professor	1 (0.8%)
<i>Experience Level</i>	
Beginner	15 (11.4%)
Intermediate	61 (46.2%)
Advanced	34 (25.8%)
Expert	22 (16.7%)
<i>Discipline Category</i>	
Engineering & Technology	87 (65.9%)
Social Sciences	42 (31.8%)
Life Sciences & Medicine	3 (2.3%)
<i>Discipline</i>	
Computer Science & Information Systems	51 (38.6%)
Library and Information Studies	30 (22.7%)
Electrical & Electronic Engineering	30 (22.7%)
Communication & Media Studies	8 (6.1%)
Mechanical, Aeronautical & Manufacturing Engineering	5 (3.8%)
Biological Sciences	2 (1.5%)
Statistics & Operational Research	1 (0.8%)

Demographic Variable	Number of Participants
Education	1 (0.8%)
Politics & International Studies	1 (0.8%)
Economics & Econometrics	1 (0.8%)
Civil & Structural Engineering	1 (0.8%)
Psychology	1 (0.8%)
<i>NTU vs non-NTU</i>	
NTU	87 (65.9%)
non-NTU	45 (34.1%)
<i>Country</i>	
Singapore	107 (81.1%)
India	4 (3%)
Malaysia	3 (2.3%)
Sri Lanka	3 (2.3%)
Pakistan	3 (2.3%)
Indonesia	2 (1.5%)
Germany	2 (1.5%)
Australia	1 (0.8%)
Iran	1 (0.8%)
Thailand	1 (0.8%)
China	1 (0.8%)
USA	1 (0.8%)
Canada	1 (0.8%)
Sweden	1 (0.8%)
Slovenia	1 (0.8%)

5.4.4 Study Procedure

During the start of the study, participants were requested to read the online user guide. The user guide was devised with the intention to help the participants in understanding the features of the Rec4LRW system along with the instructions to complete the evaluation exercise. The user guide content is provided in Appendix H.

The specific instructions for the participants were as follows. In Task 1, participants had to select a research topic from a list of 43 research topics in the task screen. The top 43 topics were shortlisted from the keywords used in the offline evaluation experiment. A free-text search box was not provided in the UI due to the limited number of papers in the corpus. Therefore, the intent was to provide a list of topics that had the most papers in the corpus. Out of the provided topics, participants used 29 topics (list of used topics provided in Table 5.6). On selection of topic, the system provided 30 recommendations. The evaluation screen was embedded at the bottom of the screen so that participants could complete the evaluation of the current task. The participants had to answer the mandatory survey questions and optional subjective feedback questions as a part of the evaluation.

Table 5.6

Research topics used by the participants

Research Topics	Number of users per topic
social media	16
machine learning	12
digital libraries	11
data mining	9

Research Topics	Number of users per topic
social networks	8
distributed systems	8
information retrieval	7
human computer interaction	6
sensor networks	6
wireless networks	5
embedded systems	5
genetic algorithm	5
computer-mediated communication	5
recommender systems	4
approximation algorithms	3
user-centered design	3
cloud computing	3
natural language processing	3
access control	3
user experience	2
static analysis	2
game theory	2
information visualization	2
human-robot interaction	2
collaborative filtering	1
interaction design	1
software architecture	1
mobile computing	1

Research Topics	Number of users per topic
eye tracking	1

Before executing Task 2, the participant had to add at least five papers in the seed basket (SB). Subsequently, the system provided 30 recommendations for this task. For Task 3, the participants were requested to add at least 30 papers in the personalized reading list (RL). The minimum paper count in RL was set to 30 as the threshold for highest number of shortlisted papers was 26 (for the article-type ‘generic research paper’). The three other article-types provided for the study were *conference full paper*, *conference poster* and *case study*. The shortlisted papers count for these article-types was fixed by taking average of the references count of the related papers from the ACM DL dataset. The participant had to then select the article-type and run the task so that the system could retrieve the shortlisted papers. At the end of this task, the evaluation questionnaire comprised of two sets of questions for the task and the overall system respectively. By the end of the study, each participant answered a total of three questionnaires as part of the evaluation.

5.4.5 Quantitative Data Analysis

5.4.5.1 Common Analysis Procedures

A five-point Likert scale was provided for measuring participant response for survey-type questions in the three questionnaires. The response values ‘Agree’ and ‘Strongly Agree’ were the two values considered for the calculation of agreement percentages for the evaluation measures. An agreement percentage above 75% was considered as an indication of higher agreement from the participants. Descriptive statistics were used to measure central tendency. Independent samples t-test was used to check the presence of statistically significant difference in the mean values of the students and

staff group, for the testing the hypothesis. Spearman correlation coefficient was used to measure the correlation between the measures. For the predictive model, multiple linear regression was used. For validating the regression models from the three tasks, 80% of the samples were selected randomly to predict the dependent variable in each task. The means of the original values and predicted value were compared using the paired samples t-test to ascertain the difference. If the mean values are very close to each other, the t-value should be close to zero and there should be no statistical significance difference between the means. If this condition is met, the model's validity can be considered to be strong. Statistical significance was set at $p < 0.05$. Statistical analyses were done using SPSS 21.0 and R.

To handle participant dropout during the study, the following approach was used. If the participant left before completing Task 1 evaluation, their data was not recorded in the system. If the participant left during Task 2 evaluation, the Task 1 survey responses were considered. Similarly, if the participant left during Task 3 evaluation, both Task 1 and 2 survey responses were considered. Hence, there was varying number of participants for the task-level evaluation. This variation did not affect the analysis of the evaluation data because although the three tasks were part of the same system, the recommendations from these tasks were generated for specific requirements.

5.4.5.2 Task 1 Evaluation

Evaluation Measures

The 14 survey questions and the corresponding measures used for the Task 1 evaluation are listed in Table 5.7. The selection of the evaluation measures were performed based on two criteria. The first criteria was to test the requirements

identified for the tasks and the second criteria was to test the characteristics of the recommendations based on previous SPRS studies. The measures *popularity*, *recency* and *diversity* correspond to the requirements of the reading list (refer Section 4.3.3.1 in Chapter 4). The *interdisciplinarity* measure corresponds to the ability of the *TPC* technique in identifying interdisciplinary papers (refer Section 4.2.3.1 in Chapter 4). The measures *good_spread*, *familiarity*, *serendipity*, *good_list* and *user_satisfaction* were adopted from an earlier study (Mcnee, 2006) in which recommendations for multiple information seeking tasks were evaluated. The measures *topical_relevance* and *usefulness* were adopted from (Ekstrand et al., 2010). The measures *good_spread*, *usefulness* and *good_list* are standard measures used in general RS studies (Knijnenburg, Willemsen, Gantner, Soncu, & Newell, 2012). The measures *good_mix* and *expansion_required* were novel measures added for this study's requirements. A screenshot of the survey questionnaire is provided in Figure 5.5.

Table 5.7

Survey questions and corresponding measures for Task 1

Sl.no	Question	Measure
1	The recommendation list is relevant to the research topic	Topical_Relevance
2	The recommendation list consists of a good spread of papers for the research topic	Good_Spread
3	The recommendation list consists of papers from different sub-topics	Diversity
4	The recommendation list consists of interdisciplinary papers	Interdisciplinarity

Sl.no	Question	Measure
5	The recommendation list consists of papers that appear to be popular papers for the research topic	Popularity
6	The recommendation list consists of a decent quantity of recent papers	Recency
7	The recommendation list consists of a good mix of diverse, recent, popular and literature survey papers	Good_Mix
8	The papers in the recommendation list appear familiar to you	Familiarity
9	The papers in the recommendation list are unknown to you	Novelty
10	The recommendation list consists of some unexpected papers that you were not expecting to see	Serendipity
11	The papers in the recommendation list are useful for reading at the start of your literature review	Usefulness
12	This is a good recommendation list, at an overall level	Good_List
13	There is a need to further expand this recommendation list	Expansion_Required
14	Please select your satisfaction level for this recommendation list	User_Satisfaction

Your username

1. The recommendation list...

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Is relevant to the research topic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consists of a good spread of papers for the research topic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consists of papers from different sub-topics	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consists of interdisciplinary papers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consists of papers that appear to be popular papers for the research topic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consists of a decent quantity of recent papers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consists of a good mix of diverse, recent, popular and literature survey papers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The papers in the recommendation list appear familiar to you	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The papers in the recommendation list are unknown to you	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The recommendation list consists of some unexpected papers that you were not expecting to see	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The papers in the recommendation list are useful for reading at the start of your literature review	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This is a good recommendation list, at an overall level	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
There is a need to further expand this recommendation list	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3. Please select your satisfaction level for this recommendation list

☐ Very Satisfied
☐ Satisfied
☐ Neutral
☐ Dissatisfied
☐ Very Dissatisfied

4. From the displayed information, what features did you like the most?

5. Please provide your personal feedback about the execution of this task

Figure 5.5. Task evaluation questionnaire for Task 1

Findings

Agreement on the Task Results

The agreement percentages for the measures by the participant group for Task 1 are illustrated in Figure 5.6. The students group had high agreement for the measures *topical_relevance* (83.87%), *good_spread* (77.42%), *diversity* (75.81%), *recency* (79.65%), *usefulness* (76.72%), *expansion_required* (82.26%) and *user_satisfaction* (75.81%). Graduate research students stand to be benefitted by this automated approach of generating reading list since quality measures indicated that the list meets their expectations. On the other hand, staff group had high agreement for only two measures *topical_relevance* (75.71%) and *usefulness* (78.97%). Even though, the recommended papers were limited to the papers in the ACM DL dataset, many participants were expecting to see some other familiar papers from information sources such as IEEE Xplore. This expectation might have hampered the evaluation

for the staff group. Secondly, some of the staff participants felt that the broad topics available for selection, was not convenient as they wanted to search for specific sub-topics in their area of expertise. Interestingly, the agreement level for the *interdisciplinarity* measure was high for the staff participants (61.43%) than students (56.45%). This observation could be attributed to the experience of researchers in identifying certain journals and conferences as being specific to their discipline than others.

From the agreement results, it is evident that the researchers found the task to be useful in retrieving relevant, recent and diverse set of papers, thereby corroborating the results of the offline evaluation experiment. Both the groups felt that the paper count of 30 was not sufficient for a reading list (82.26% for students, 71.43% for staff). They wanted the system to allow the user to set the number of recommendations.

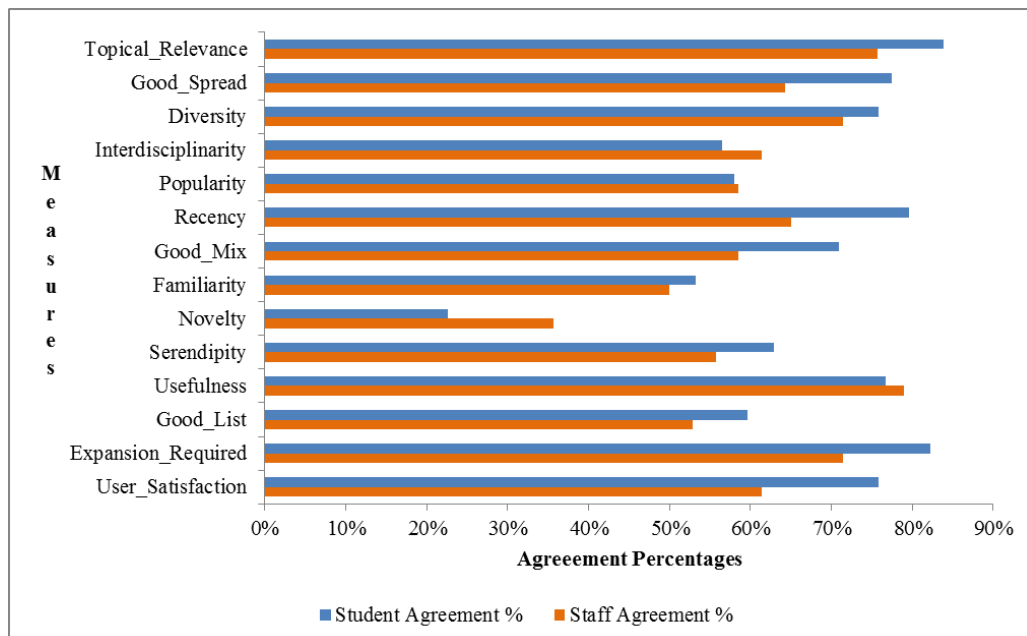


Figure 5.6. Agreement percentages for the evaluation measures of Task 1

Hypothesis Testing

The results of the independent samples t-test are provided in Table 5.8. The difference between the groups was statistically significant for five measures *topical_relevance*, *good_mix*, *novelty*, *good_list* and *user_satisfaction*. Hence, the hypothesis was met for these five measures. Since three of these five measures are overall output quality measures, a claim can be made that students clearly preferred the recommendations more than staff. The difference in *topical_relevance* measure could be attributed to the expectations of the participant. Broad input topics were an issue for staff and they expected the system to retrieve papers very much relevant to their usage context which they had in mind. In contrast, students evaluated the recommended papers for closeness to the input topic, hence they evaluated more favorably.

Among the 14 evaluation measures, *good_spread*, *good_mix*, *good_list* and *user_satisfaction* are the overall quality measures. Since the students rated highly for the measures *good_mix* and *good_list*, there appears to be a natural propensity to be more satisfied with the results. At an overall level, the measures with the highest mean values were *topical_relevance* ($M=4.13$ for students, $M=3.84$ for staff), *good_spread* ($M=3.90$ for students, $M=3.71$ for staff), *diversity* ($M=3.94$ for students, $M=3.80$ for staff), *good mix* ($M=3.82$ for students, $M=3.54$ for staff) and *user_satisfaction* ($M=3.76$ for students, $M=3.51$ for staff).

Table 5.8

Independent samples t-test results of Task 1

Measure	t value	Students	Staff
		<i>M</i> (SD)	<i>M</i> (SD)
Topical_Relevance	2.201*	4.13 (0.713)	3.84 (0.773)

Measure	t value	Students	Staff
		<i>M</i> (SD)	<i>M</i> (SD)
Good_Spread	1.401	3.90 (0.783)	3.71 (0.764)
Diversity	1.067	3.94 (0.744)	3.80 (0.714)
Interdisciplinarity	0.254	3.58 (0.860)	3.54 (0.846)
Popularity	0.638	3.61 (0.837)	3.51 (0.928)
Recency	1.296	3.65 (0.960)	3.43 (0.957)
Good_Mix	1.923*	3.82 (0.800)	3.54 (0.863)
Familiarity	0.828	3.42 (0.860)	3.29 (0.980)
Novelty	-1.676*	2.74 (0.904)	3.01 (0.955)
Serendipity	-0.455	3.48 (0.882)	3.56 (0.958)
Usefulness	0.985	3.74 (0.700)	3.60 (0.923)
Good_List	1.912*	3.66 (0.745)	3.37 (0.966)
Expansion_Required	1.149	3.97 (0.677)	3.80 (0.957)
User_Satisfaction	1.818	3.76 (0.619)	3.51 (0.880)

* $p < .05$

Correlation and Regression Analysis

The measure combinations with high degree of correlation (greater than 0.5) are displayed in in Table 5.9. The *popularity* measure was found to be highly correlated with three output quality measures *usefulness* ($R=0.62$), *good_list* ($R=0.57$) and *user_satisfaction* ($R=0.58$). This finding indicates participants' expectation of finding popular papers in the list. If more popular/seminal papers were found in the list, the participants' satisfaction with the reading list increased. The other correlation results are on expected lines with *usefulness* having positive correlations with *good_list*

($R=0.67$) and *user_satisfaction* ($R=0.65$), indicating that quality measures tend to be rated similarly.

Table 5.9

Spearman correlation between measures of Task 1

Measure	Measure	R (95% CI)
	Usefulness	0.62
Popularity	Good_List	0.57
	User_Satisfaction	0.58
Usefulness	Good_List	0.67
	User_Satisfaction	0.65
Good_List	User_Satisfaction	0.72

Results from multiple linear regression testing are displayed in Table 5.10. The model was built with *good_list* as the dependent variable and the other 13 measures as the independent variables. The multiple correlation coefficient R value of 0.85 and the adjusted R^2 value of 0.72 indicate decent level of prediction at a statistically significant level. The model fit could potentially improve with more participants. Five independent variables *recency*, *novelty*, *serendipity*, *usefulness* and *user_satisfaction* were found to be statistically significant predictors in the model. The predictive ability of *usefulness* and *user_satisfaction* is understandable as participants who perceive the papers to be useful for their LR, will most probably be satisfied.

Such participants are predisposed to consider the list as a good recommendations list. However, the high estimates of the other three measures are an interesting case. The reading list search task is the first search task in the LR process where researchers are on the lookout for papers for acquiring an initial understanding

of the particular research area. Even though, seminal/popular papers are important papers for LR, recent and new papers are equally important as researchers would prefer to learn about unknown research studies. Therefore, the predictive ability of the three measures *recency*, *novelty* and *serendipity* is ratified. This observation is rather unique to this study as earlier studies have largely ignored the recommendations of recent and serendipitous papers. When the model was tested for validation using the paired samples t-test, the *t* value was very low ($t=-0.062$) and there was no statistically significant difference between the original and predicted values. This finding shows that model can be considered to be strongly valid.

Table 5.10

Multiple linear regression results of Task 1

	Estimate	SE	t value	<i>p</i>
Intercept	.132	.456	.289	.773
Topical_Relevance	.059	.069	.846	.399
Good_Spread	.077	.068	1.142	.256
Diversity	.074	.070	1.054	.294
Interdisciplinarity	.097	.062	1.562	.121
Popularity	.009	.070	.135	.893
Recency*	.149	.053	2.780	.006
Good_Mix	.041	.067	.611	.542
Familiarity	-.117	.062	-1.874	.063
Novelty*	-.122	.057	-2.128	.035
Serendipity*	-.103	.048	-2.140	.034
Usefulness*	.267	.079	3.388	.001

	Estimate	SE	t value	p
Expansion_Required	-.010	.055	-.180	.857
User_Satisfaction*	.461	.093	4.968	.000
Residual standard error: 0.487 on 118 <i>df</i>				
R: 0.850 Multiple R ² : 0.723, Adjusted R ² : 0.693				
<i>F</i> -statistic: 23.701 on 13 and 118 <i>df</i>				
Model validation with paired samples t-test: <i>t</i> : -0.062 (<i>p</i> =0.951), <i>r</i> : 0.861 (<i>p</i> =0.000)				

5.4.5.3 Task 2 Evaluation

Evaluation Measures

The 16 survey questions and the corresponding evaluation measures used in the Task 2 are provided in Table 5.11. *Seedbasket_similarity*, *shared_corelations* and *seedbasket_usefulness* were feature-related measures meant for evaluating the features provided as part of this task. The remaining 13 measures have been retained from Task 1 evaluation. *Seedbasket_similarity* and *shared_corelations* measures were included for testing the first and third requirements for Task (refer Section 4.3.4.1 in Chapter 3) while the second requirement was tested by the measures *diversity* and *interdisciplinarity*. A screenshot of the survey questionnaire is provided in Figure 5.7.

Table 5.11

Evaluation questions and corresponding measures of Task 2

Sl.no	Question	Measure
1	The recommendation list consists of papers that are similar to the papers in the seed basket	Seedbasket_Similarit y
2	The recommendation list consists of papers that have	Shared_Corelations

Sl.no	Question	Measure
	shared co-references and co-citations with the papers in the seed basket	
3	The recommendation list is relevant to the research topic	Topical_Relevance
4	The recommendation list consists of a good spread of papers for the research topic	Good_Spread
5	The recommendation list consists of papers from different sub-topics	Diversity
6	The recommendation list consists of interdisciplinary papers	Interdisciplinarity
7	The recommendation list consists of papers that appear to be popular papers for the research topic	Popularity
8	The recommendation list consists of a decent quantity of recent papers	Recency
9	The papers in the recommendation list appear familiar to you	Familiarity
10	The papers in the recommendation list are unknown to you	Novelty
11	The recommendation list consists of some unexpected papers that you were not expecting to see	Serendipity
12	The papers in the recommendation list are useful for reading at the start of your literature review	Usefulness
13	This is a good recommendation list, at an overall level	Good_List
14	There is a need to further expand this recommendation	Expansion_Required

Sl.no	Question	Measure
	list	
15	Please select your satisfaction level for this recommendation list	User_Satisfaction
16	The feature of adding papers to the seed basket to generate similar paper recommendations is a useful feature	Seedbasket_Usefulne SS

[Click here to start evaluation of this task](#)

- The recommendation list...

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Consists of papers that are similar to the papers in the seed basket	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consists of papers that have shared co-references and co-citations with the papers in the seed basket	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
- The recommendation list...

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Is relevant to the research topic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consists of a good spread of papers for the research topic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consists of papers from different sub-topics	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consists of interdisciplinary papers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consists of papers that appear to be popular papers for the research topic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Consists of a decent quantity of recent papers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
- | | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| The papers in the recommendation list appear familiar to you | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| The papers in the recommendation list are unknown to you | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| The recommendation list consists of some unexpected papers that you were not expecting to see | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| The papers in the recommendation list are useful for reading during your literature review | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| This is a good recommendation list, at an overall level | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| There is a need to further expand this recommendation list | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
- Please select your satisfaction level for this recommendation list
 - ☐ Very Satisfied
 - ☐ Satisfied
 - ☐ Neutral
 - ☐ Dissatisfied
 - ☐ Very Dissatisfied
- The feature of adding papers to the seed basket to generate similar paper recommendations is a useful feature
 - ☐ Strongly Agree
 - ☐ Agree
 - ☐ Disagree
 - ☐ Strongly Disagree
 - ☐ Not Sure

Figure 5.7. Evaluation screen of the similar papers task (Task 2)

Findings

Agreement on the Task Results

In Figure 5.8, the agreement percentages for the 16 measures are displayed for the two groups. The measures that met the threshold criteria are *seedbasket_similarity* (88.71% for students, 75.41% for staff), *shared_correlations* (82.26% for students), *topical_relevance* (90.32% for students, 83.61% for staff), *good_spread* (87.10% for students), *diversity* (79.03% for students), *usefulness* (82.26% for students, 78.69% for staff), *good_list* (79.03% for students), *seedbasket_usefulness* (96.77% for students, 95.08% for staff). The high agreement on the feature-related measures indicate high topical similarity with the seed basket (SB) papers since the IDSP technique's design considers all the SB papers for formulating recommendations. The ability of the technique in covering a wide variety of sub-topics in the research area is vindicated with *good_spread* and *diversity* measures. This finding is partially attributed to the ability of the IDSP technique's topical similarity module in finding papers that are not part of the citation networks of the SB papers. The agreement level of the measures *usefulness* and *good_list* indicate to corresponding higher levels of satisfaction. However, the *user_satisfaction* percentage of both groups was around a decent range (approximately 70%) which is below the set threshold for this study. This finding can be linked to the measure *expansion_required* (74.35% for students, 72.35% for staff) which highlights the expectation of participants for acquiring more papers in the list, even though the responses for the other measures were favorable.

The measures *novelty* (27.42% for students, 36.07% for staff) and *serendipity* (54.84% for students, 57.38% for staff) had low agreement percentages. The ACM DL dataset used in the study consists of papers published before 2011. Therefore, the recommended papers would have been largely familiar to participants. Serendipitous

discovery of new papers is also affected by the dataset. Secondly, the current design of the IDSP technique is not prioritized for finding serendipitous papers. The results of the other measures indicate that participants were largely in favor with the recommended papers.

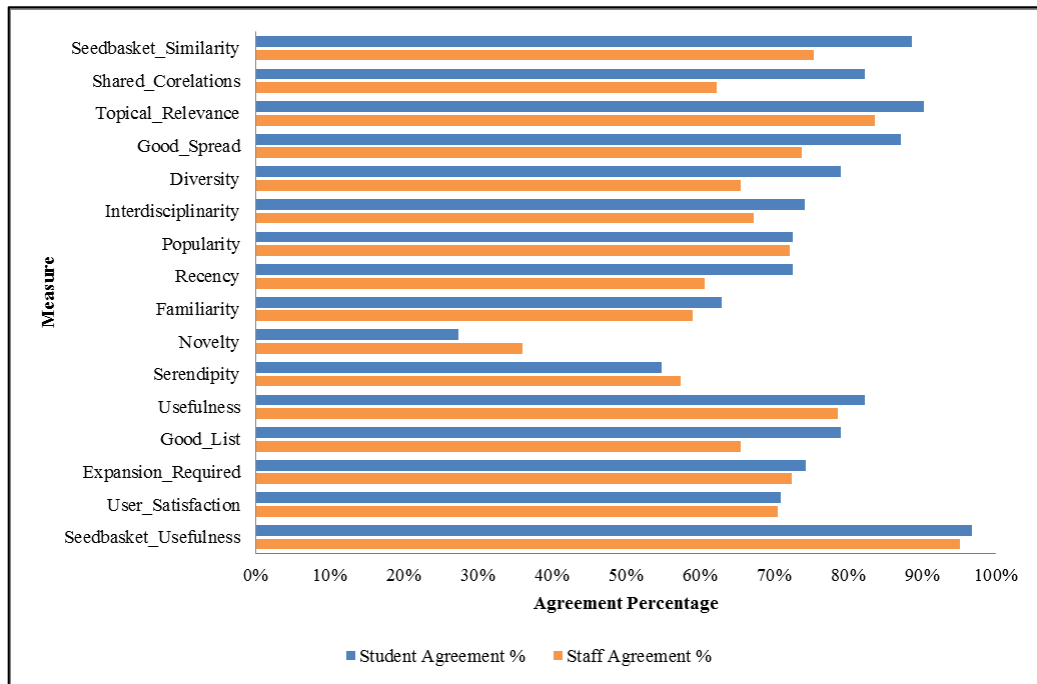


Figure 5.8. Agreement percentages of the evaluation measures of Task 2

Hypothesis Testing

Table 5.12 lists the independent samples t-test results for the two groups. As for the hypothesis, students group rated higher than staff group at a statistically significant level for five measures *seedbasket_similarity*, *shared_correlations*, *topical_relevance*, *good_spread* and *good_list*. The difference for the measures *seedbasket_similarity* ($M=4.02$ for students, $M=3.77$ for staff) and *shared_correlations* ($M=3.90$ for students, $M=3.61$ for staff) is an interesting case as these are easily inferable measures. The participants have to merely identify whether the recommended papers are similar to the SB papers. It could be speculated that students were more confident of the

similarity than staff. In addition, experienced researchers have higher awareness of their expertise areas; therefore they would have expected more closely-related papers in the list or noticed papers with weak relations to the SB papers. The differences for the measures *topical_relevance* ($M=4.13$ for students, $M=3.84$ for staff), *good_spread* ($M=3.98$ for students, $M=3.69$ for staff) and *good_list* ($M=3.85$ for students, $M=3.61$ for staff) can be explained by the corresponding higher ratings for other measures by the students group.

The lack of statistically significant differences for the other measures indicates two observable characteristics of the recommended papers list. The heterogeneity of the list in providing different types of paper (recent, paper, diverse and interdisciplinary) is acknowledged by the participants. Secondly, the findings vindicate the nature of the task in improving the discovery of relevant papers from the first task (reading list task). Therefore there is a semblance of uniformity in participants' evaluation responses, with an inclination towards higher agreeability.

Table 5.12

Independent samples t-test results of Task 2

Measure	t	Students	Staff
		<i>M</i> (SD)	<i>M</i> (SD)
Seedbasket_Similarity	1.699*	4.02 (0.558)	3.77 (0.99)
Shared_Corelations	2.049*	3.90 (0.620)	3.61 (0.954)
Topical_Relevance	2.282*	4.13 (0.558)	3.84 (0.840)
Good_Spread	2.497*	3.98 (0.496)	3.69 (0.786)
Diversity	1.250	3.87 (0.640)	3.70 (0.823)
Interdisciplinarity	0.154	3.76 (0.645)	3.74 (0.814)

Measure	t	Students	Staff
		<i>M</i> (SD)	<i>M</i> (SD)
Popularity	0.135	3.85 (0.721)	3.84 (0.820)
Recency	0.554	3.69 (0.934)	3.61 (0.802)
Familiarity	0.840	3.60 (0.877)	3.46 (0.941)
Novelty	-1.356	2.81 (0.884)	3.03 (0.966)
Serendipity	0.055	3.47 (0.824)	3.46 (0.923)
Usefulness	0.820	3.97 (0.677)	3.85 (0.872)
Good_List	1.861*	3.85 (0.568)	3.61 (0.881)
Expansion_Required	0.949	3.71 (0.755)	3.56 (1.009)
User_Satisfaction	1.267	3.81 (0.649)	3.62 (0.934)
Seedbasket_Usefulness	0.390	4.32 (0.790)	4.26 (0.730)

* indicates $p < 0.05$

Correlation and Regression Analysis

In Table 5.13, the measure combinations with correlation coefficient values above the threshold value of ‘0.5’ are displayed. The moderate correlation between *seedbasket_similarity* and *shared_corelations* ($R=0.502$) is an expected observation as both the measures point to the same aspect of topical similarity between the recommended papers and SB papers. The correlation between *shared_corelations* and *good_spread* ($R=0.566$) is interesting since they are conceptually disparate features. The validity of this finding is to be established in future studies. The correlation of the measure *good_list* with *good_spread* ($R=0.503$) and *usefulness* ($R=0.569$) is another expected finding as they are output quality measures. These three output quality measures are in turn positively correlated with the fourth and important output quality measure *user_satisfaction*. Hence, there is consistency among output quality measures.

The measures *topical_relevance* (R=0.633) and *familiarity* (R=0.539) are also correlated with *user_satisfaction*. The inference is that if participants find known relevant papers for the given research topic, they tend to more satisfied with the overall list. It is to be noted that *familiarity* is a measure with little use in a real world setting as the recommended papers are supposed to be new to the user while searching for papers for an unknown research topic.

Table 5.13

Measure combinations with moderate to high correlations of Task 2

Measure 1	Measure 2	R (95% CI)
Seedbasket_Similarity	Shared_Corelations	0.502
Shared_Corelations	Good_Spread	0.566
Good_Spread	Good_List	0.503
Usefulness		0.569
Topical_Relevance		0.633
Good_Spread		0.525
Familiarity	User_Satisfaction	0.539
Usefulness		0.608
Good_List		0.607

Results from multiple linear regression testing are displayed in Table 5.14. The model was built with *good_list* as the dependent variable and the 14 other measures as the independent variables. The multiple correlation coefficient R value of 0.70 and the adjusted R² value of 0.50 indicate moderate level of prediction at a statistically significant level. Only two independent variables *seedbasket_similarity* and *usefulness* were found to be statistically significant predictors in the model, unlike Task 1 which

had five predictors. *Usefulness* was the main predictor for agreeability on a good list for this task. *Usefulness* has remained an important measure from earlier studies (Ekstrand et al., 2010; Knijnenburg et al., 2012), irrespective of the domain. Even though, recent SPRS studies have made cases for considering new measures such as *diversity* (Küçükünç et al., 2015) and *serendipity* (Sugiyama & Kan, 2011), the importance of traditional measures such as *usefulness* cannot be understated. The predictive ability of *seedbasket_similarity* is validated by the high agreement percentages from both the groups (88.71% for students, 75.41% for staff). When the model was tested for validation using the paired samples t-test, the *t* value was low ($t=0.603$) and there was no statistically significant difference between the original and predicted values. This finding shows that model can be considered to be moderately valid when compared to the Task 1 model.

The conclusion from this test result is that the recommendation technique will be considered useful if the recommendations are similar to the input papers (seed basket papers). In the Rec4LRW system, this relationship was outwardly made apparent to the participants through the shared co-relations features in the user-interface (refer Section 4.4.3.1 from Chapter 4). This task's IDSP technique and shared co-relations display features complement each other, thereby reiterating the usefulness of the threefold intervention framework.

Table 5.14

Multiple linear regression results of Task 2

	Estimate	SE	t value	<i>p</i>
Intercept	0.292	0.559	0.523	0.602

	Estimate	SE	t value	<i>p</i>
Seedbasket_Similarity*	0.224	0.088	2.539	0.013
Shared_Corelations	-0.038	0.090	-0.421	0.674
Topical_Relevance	0.060	0.122	0.496	0.621
Good_Spread	0.171	0.104	1.646	0.103
Diversity	0.055	0.084	0.655	0.514
Interdisciplinarity	0.022	0.081	0.272	0.786
Popularity	-0.019	0.084	-0.229	0.819
Recency	-0.021	0.076	-0.278	0.782
Familiarity	0.043	0.078	0.554	0.581
Novelty	0.082	0.066	1.247	0.215
Serendipity	-0.003	0.063	-0.051	0.959
Usefulness*	0.223	0.096	2.319	0.022
Expansion_Required	-0.089	0.063	-1.424	0.157
User_Satisfaction	0.200	0.113	1.769	0.080

Residual standard error: 0.561 on 108 *df*

R: 0.708 Multiple R²: 0.501, Adjusted R²: 0.436

F-statistic: 7.750 on 14 and 108 *df*

Model validation with paired samples t-test: *t*: 0.603 (*p*=0.548), *r*: 0.680 (*p*=0.000)

* indicates *p* < 0.05

5.4.5.4 Task 3 Evaluation

Evaluation Measures

Majority of the measures from the first two tasks were not applicable for this task due to its scope. For this task evaluation, the measures were selected based on the key aspects of the task. The seven survey questions and the corresponding measures are provided in Table 5.15. The measures *relevance*, *usefulness*, *importance*, *certainty*, *good_list* and *improvement_needed* were meant to ascertain the quality of the recommendations. The *importance* measure was selected for testing the main characteristic of this task's CNS technique. The corresponding requirement is listed in Sections 4.3.5.1 of Chapter 4. The *certainty* measure was selected as it is essential to understand whether the participants have a strong inclination to cite the recommended papers. The final measure *shortlisting_feature* was used to identify whether participants would be interested to use this task in current academic search systems and digital libraries. The *relevance* has a modified definition for this task, in the contrast to the earlier two tasks. This modified version was added for testing the second requirement for Task 3. A screenshot of the survey questionnaire is provided in Figure 5.9.

Table 5.15

Evaluation questions and corresponding measures of Task 3

Sl.no	Question	Measure
1	The shortlisted papers are relevant to my article-type preference	Relevance
2	The shortlisted papers are useful for inclusion in my	Usefulness

Sl.no	Question	Measure
	manuscript	
3	The shortlisted papers comprises of important papers from my reading list	Importance
4	The shortlisted list comprises of papers which I would definitely cite in my manuscript	Certainty
5	This is a good recommendation list, at an overall level	Good_List
6	There is a need to further improve this shortlisted papers list	Improvement_Needed
7	I would like to see the feature of shortlisting papers from reading list based on article-type preference, in academic search systems and databases	Shortlisting_Feature

Your username

1.

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
The shortlisted papers are relevant to my article-type preference	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The shortlisted papers are useful for inclusion in my manuscript	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The shortlisted papers comprises of important papers from my reading list	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The shortlisted list comprises of papers which I would definitely cite in my manuscript	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
This is a good shortlisted papers list, at an overall level	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
There is a need to further improve this shortlisted papers list	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2.

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
I would like to see the feature of shortlisting papers from reading list based on article-type preference, in academic search systems and databases	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would like to see the feature of managing reading list and seed basket papers between the three tasks in academic search systems and databases	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3. From the displayed information, what features did you like the most?

4. Please provide feedback on the execution of this task

Figure 5.9. Evaluation screen of the shortlisting papers task (Task 3)

Findings

Agreement on the Task Results

The agreement percentages for the seven measures by the participant groups are shown in Figure 5.10. As observed from findings of the previous tasks, the agreement percentages of students was consistently higher than the staff with the biggest difference found for the measures *usefulness* (82.00% for students, 64.15% for staff) and *good_list* (76.00% for students, 62.26% for staff). The quality measures *importance* (85.96% for students, 77.97% for staff) and *shortlisting_feature* (84.21% for students, 74.58% for staff) had the highest percentages at an overall level. This observation validates the usefulness of the technique in identifying popular/seminal papers from the reading list. Due to favorable percentages for the most measures, the lowest agreement values were observed for the measure *improvement_needed* (57.89% for students, 57.63% for staff). The results for the measure *certainty* (70% for students, 62.26% for staff) indicate some level of reluctance among the participants in being confident of citing the papers. Citation of a particular paper is subject to the particular citation context in the manuscript, therefore not all participants would be able to prejudge their citation behavior. In summary, participants seem to acknowledge the usefulness of the task in identifying important papers from the reading list. However, there is an understandable lack of inclination in citing these papers. This issue is to be addressed in future versions of the Rec4LRW.

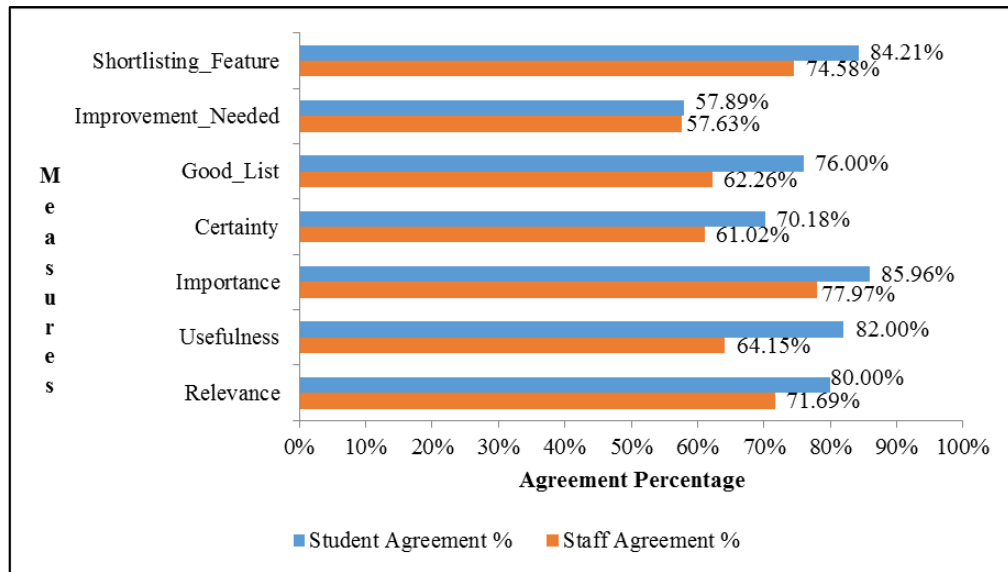


Figure 5.10. Agreement percentages of the evaluation measures of Task 3

Hypothesis Testing

Table 5.16 lists the independent samples t-test results for the two groups. There were no statistically significant difference between the two groups for any of the seven measures at $p < 0.05$ level. Therefore, the hypothesis was not met for this task, a finding different from the earlier tasks. Noticeably, standard deviations (SD) were high for the staff group for most of the measures. The measures *usefulness* and *good_list* have high SDs ($SD > 1$) indicating differences in perceptions among the staff participants. The mean values of students were consistently higher than staff with the exception of two measures *improvement_needed* ($M=3.63$ for students, $M=3.75$ for staff) and *shortlisting_feature* ($M=4.00$ for students, $M=4.05$ for staff). In the case of the former, the difference can be attributed to the higher expectation of staff in terms of the task recommendations. For the latter case, this difference can be considered as a good indication as participants with higher experience levels seem to prefer this type of task. Clearly, they would want this type of feature included in digital libraries and also in reference management systems.

Table 5.16

Independent samples t-test results of Task 3

Measure	t	Students	Staff
		M (SD)	M (SD)
Relevance	0.147	3.93 (0.728)	3.90 (0.912)
Usefulness	0.961	3.89 (0.646)	3.70 (1.081)
Importance	0.737	3.91 (0.510)	3.80 (0.768)
Certainty	1.598	3.72 (0.675)	3.40 (0.995)
Good_List	0.444	3.89 (0.724)	3.80 (1.056)
Improvement_Needed	-0.523	3.63 (0.837)	3.75 (0.967)
Shortlisting_Feature	-0.267	4.00 (0.681)	4.05 (0.826)

Correlation and Regression Analysis

In Table 5.17, the measure combinations with correlation coefficient values above the threshold value of '0.5' are listed for Task 3. The correlations were in the moderate range for all combinations with highest values for the combinations *usefulness* and *good_list* (R=0.62), followed by *relevance* and *good_list* (R=0.602). These three measures had also comparatively high percentages than other measures (refer Figure 5.10). The interpretation of these results is that if the participants find the shortlisted papers relevant to their article-type preference and useful for inclusion in their manuscript drafts, the shortlisted papers will be perceived as an adequate recommendations list.

Table 5.17

Measure combinations with moderate to high correlations of Task 3

Measure 1	Measure 2	R (95% CI)
Relevance	Usefulness	0.587
	Importance	0.554
	Good_List	0.602
Usefulness	Certainty	0.503
	Good_List	0.62
Importance	Good_List	0.501

Results from multiple linear regression testing are displayed in Table 5.18. The model was built with *good_list* as the dependent variable and the six other measures as the independent variables. The multiple correlation coefficient R value of 0.72 and the adjusted R^2 value of 0.50 indicate moderate level of prediction at a statistically significant level. Three independent variables *relevance*, *usefulness* and *certainty* were found to be statistically significant predictors in the model. Out of the three independent variables, *usefulness* had the highest estimate in the model, thereby validating the strong correlation between the two measures (refer Table 5.17). This is an interesting finding since *certainty* is expected to be the most positive measure from the list. The inference here is that if the participants feel that the shortlisted papers are useful for citation in their manuscripts, they would perceive the list to be good for their requirement. When the model was tested for validation using the paired samples t-test, the t value was low ($t=0.647$) and there was no statistically significant difference between the original and predicted values. This finding shows that model can be considered to be moderately valid.

Table 5.18

Multiple linear regression results of Task 3

	Estimate	SE	t value	<i>p</i>
Intercept	.372	.498	.748	.456
Relevance	.212*	.093	2.280	.025
Usefulness	.395*	.097	4.051	.000
Importance	.114	.122	.936	.351
Certainty	.221*	.094	2.351	.021
Improvement_Needed	-.041	.067	-.619	.537
Shortlisting_Feature	-.008	.075	-.101	.920
Residual standard error: 0.592 on 109 <i>df</i>				
Multiple R^2 : 0.722, Adjusted R^2 : 0.495				
<i>F</i> -statistic: 19.820 on 6 and 109 <i>df</i>				
Model validation with paired samples t-test: <i>t</i> : 0.647 (<i>p</i> =0.519), <i>r</i> : 0.744 (<i>p</i> =0.000)				

5.4.5.5 System-level Evaluation***Evaluation Constructs and Measures***

For the system level evaluation, three constructs were shortlisted from prior studies.

The constructs *Effort to use the System* and *Perceived System Effectiveness* were adopted from a user experience RS study (Knijnenburg et al., 2012) where evaluation constructs were validated in a series of user studies. The former construct measures the user effort through five measures while latter measures the effectiveness through six measures. The third construct *Perceived Usefulness* was adopted from the TAM model (Venkatesh & Bala, 2008) and it comprises of six internal measures. The three

constructs and the constituent measures used for the system-level evaluation are listed in Table 5.19.

Table 5.19

System-level evaluation constructs and constituent measures

	Construct	Measure	Question
1	Effort to use the System	Convenience	The system is convenient
2		Effort_Required	I have to invest a lot of effort in the
3			system
4		Mouse_Clicks	It takes many mouse-clicks to use the
5		Little_Time	system
6	Perceived System Effectiveness	Much_Time	Using the system takes little time
7			It takes too much time before the
8		Recommend	system provides adequate
9			recommendations
10		Pleasant_Experience	I would recommend the system to
11		Useless	others
		Awareness	Using the system is a pleasant
		Better_Choice	experience
		Findability	The system is useless
			The system makes me more aware of
			my choice options
			I make better choices with the system
			I can find better papers by using the

	Construct	Measure	Question
			system
12	Perceived Usefulness	Accomplish_Tasks	Using the system would enable me to accomplish tasks more quickly
13		Performance_Improvement	Using the system would improve my work performance
14		Productivity_Improvement	Using the system would improve my productivity
15		Enhance_Effectiveness	Using the system would enhance my effectiveness on the work
16		Ease_Job	Using the system would make it easier to do my job
17		Work_Usefulness	I would find the system useful in my work

Findings

Agreement on the System Aspects

Effort to use the System.

The agreement percentages of the six measures under the construct *Effort to use the System* are illustrated in Figure 5.11. For all the six measures, there was not much difference between the groups, indicating a uniform experience. High percentages for the *convenience* measure (86% for students, 84.91% for staff) indicate that the system was convenient to use for the three tasks, thereby validating the approach of utilizing the three interventions for the issue in hand. The low percentages for the measures

effort_required, *mouse_clicks* and *much_time* indicate a comfortable experience pattern for the participants. Participants did not have to expend lot of effort in executing and traversing between the tasks. In terms of time spent for the tasks execution, the moderate percentages for the measure *little_time* (70% for students, 64.15% for staff) indicate that more improvements are required for expediting the task execution in the system.

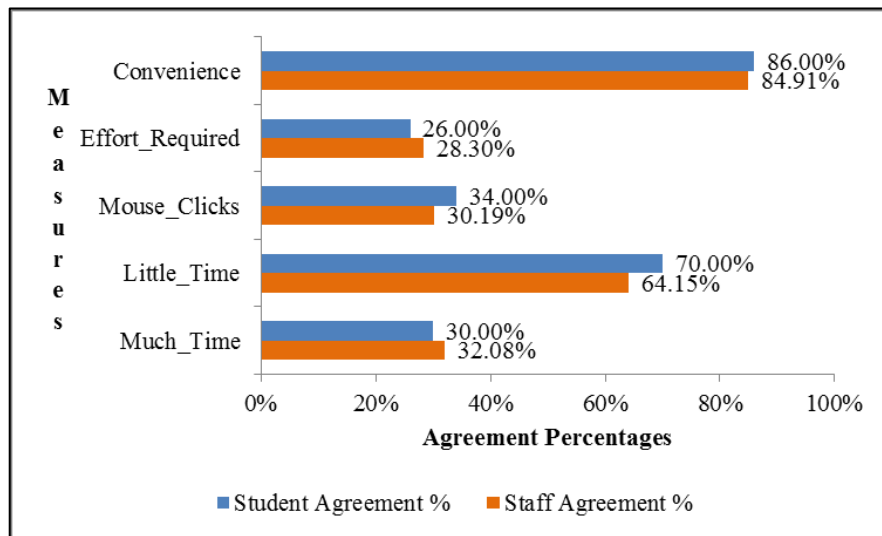


Figure 5.11. Agreement percentage results for Effort to use the System construct

Perceived System Effectiveness.

The agreement percentages of the six measures under the construct *Perceived System Effectiveness* are illustrated in Figure 5.12. The measures with the highest percentages were *pleasant_experience* (84% for students) and *awareness* (88% for students), indicating most of the students gained considerably from the system, even though the percentage for the measure *recommend* (72%) wasn't high. Interestingly, there was a huge difference with the staff for the *awareness* measure (67.92%). This difference can be attributed to the limited number of recent papers⁹ in the corpus.

⁹ The papers in the ACM dataset include papers published until 2011

Correspondingly, the percentages for the measure *findability* (64% for students, 56.60% for staff) were at a moderate level. Very low percentages for the measure *useless* (10% for students, 7.55% for staff) are a favorable outcome for the system's outlook and the future versions.

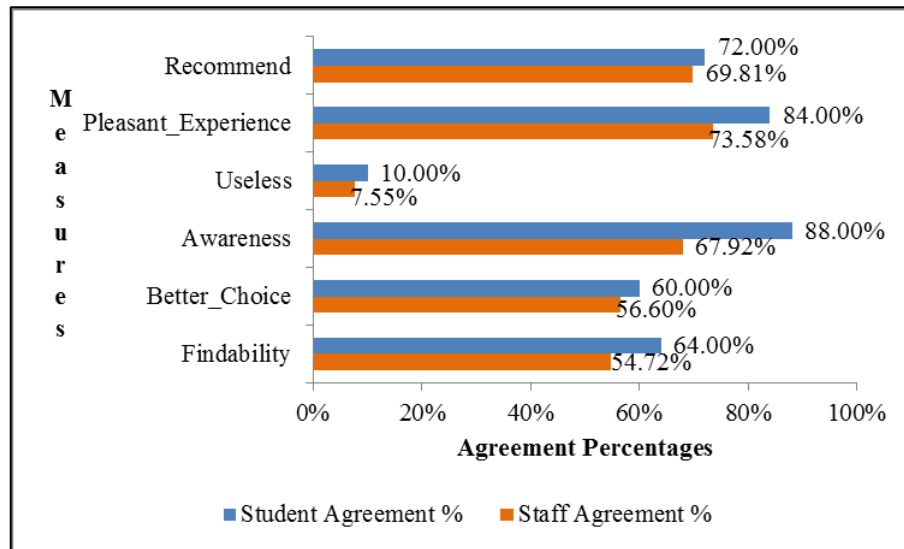


Figure 5.12. Agreement percentage results for Perceived System Effectiveness construct

Perceived Usefulness.

The agreement percentages of the six measures under the construct *Perceived Usefulness* are illustrated in Figure 5.13. This construct clearly differentiates the two groups for the constituent measures. Among the students, the percentages were high (>75%) for all measures with the exception of *productivity_improvement*. In a contrast manner, none of the measures had high percentages for the staff group. However, staff had a positive outlook for the measure *work_usefulness* (73.58%) indicating that they were open to use the system for their work, even though the system did not meet most of their preconceived requirements. Overall, the findings from this construct indicate that the system has met its objective with the students group.

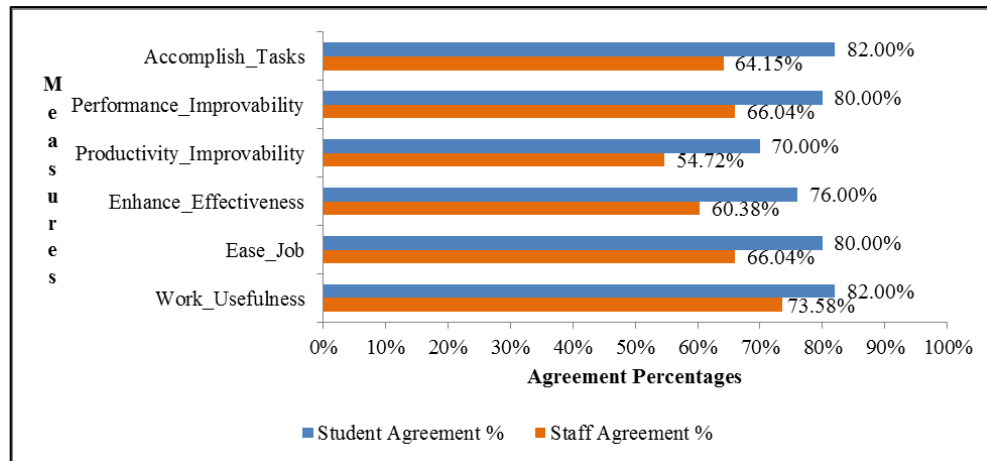


Figure 5.13. Agreement percentage results for Perceived Usefulness construct

Hypothesis Testing

Results of the independent samples t-test are presented in Table 5.20, along with the Cronbach alpha values of the three system constructs. For 12 of the 17 measures under the three constructs, there was no statistically significant difference between the two groups. The mean values for the five measures under the *Effort to use the System* construct were in similar range. These values indicate the same level of effort required from the participants, regardless of the group. Under the *Perceived System Effectiveness* construct, there was significant difference for the *awareness* measure, which can be attributed to the higher experience level of the staff group. Students, on the other hand, are less aware of the potentials papers that are to be read during LR. There was significant difference for four out of the six measures under *Perceived Usefulness*. This finding provides a strong evidence of the usability of the system for students as hypothesized, since the difference was the largest for the measures *performance_improvability* (M=3.91) and *ease_job* (M=3.96). Based on the findings from the agreement percentages and the t-test comparisons, the perception of the system appears to be consistent across the two groups for the effort involved in using

the system and its effectiveness. However, the system in its current release was found more useful for graduate students in their research work.

Table 5.20

Independent samples t-test results of system-level evaluation

Construct	α	Question	Measure	t	M (SD)	M (SD)	p
Effort to use the System	0.6	The system is convenient	Convenience	0.24	3.98	3.95	0.40
					(0.719)	(0.775)	6
		I have to invest a lot of effort in the system	Effort_Required	-0.326	2.68	2.75	0.37
					(0.909)	(1.108)	2
		It takes many mouse-clicks to use the system	Mouse_Clicks	0.282	2.75	2.69	0.38
					(1.138)	(1.133)	9
		Using the system takes little time	Little_Time	0.009	3.54	3.54	0.49
					(0.867)	(0.916)	6
		It takes too much time before the system provides adequate recommendations	Much_Time	-0.536	2.67	2.78	0.29
					(1.041)	(1.219)	6
Perceived	0.8	I would recommend the system to others	Recommend	-0.057	3.74	3.75	0.47

Construct	α	Question	Measure	t	M (SD)	M (SD)	p
System Effectiveness	1				(0.745)	(0.921)	7
		Using the system is a pleasant experience	Pleasant_Experience	0.196	3.86 (0.743)	3.83 (0.854)	0.42 3
		The system is useless	Useless	-0.294	2.07 (0.884)	2.12 (0.892)	0.38 5
		The system makes me more aware of my choice options	Awareness	2.433	3.96 (0.597)	3.64 (0.804)	0.00 8
		I make better choices with the system	Better_Choice	1.426	3.72 (0.675)	3.51 (0.898)	0.07 8
		I can find better papers by using the system	Findability	0.787	3.65 (0.79)	3.53 (0.897)	0.21 7
Perceived Usefulness	0.95	Using the system would enable me to accomplish tasks more quickly	Accomplish_Tasks	1.665	3.89 (0.646)	3.64 (0.943)	0.04 9
		Using the system would improve my work	Performance_Improvement	1.954	3.91	3.63	0.02

Construct	α	Question	Measure	t	M (SD)	M (SD)	p
		performance	ty		(0.635)	(0.908)	7
		Using the system would improve my productivity	Productivity_Improvement	1.84	3.77 (0.756)	3.49 (0.878)	0.03 4
		Using the system would enhance my effectiveness on the work	Enhance_Effectiveness	1.063	3.74 (0.669)	3.58 (0.932)	0.14 5
		Using the system would make it easier to do my job	Ease_Job	2.253	3.96 (0.706)	3.61 (0.965)	0.01 3
		I would find the system useful in my work	Work_Usefulness	1.399	3.96 (0.801)	3.75 (0.883)	0.08 2

5.4.5.6 Consolidated Findings

Agreement Percentages and Hypothesis Testing

The evaluation measures with the agreement percentages above the threshold value (75%) for the two participant groups are listed in Table 5.21, along with measures which met the hypothesis. Staff participants' experience appeared to be uniform in Tasks 1 and 2 since the *relevance* and *usefulness* measures satisfied the threshold criteria. These two measures are important measures since the former determines the recommended papers' topicality match with the input research topic while the latter determines whether participants could make use of the recommended papers.

In the hypothesis testing, Tasks 1 and 2 differentiated the two groups for at least four measures each. Interestingly, none of the measures met the hypothesis for Task 3 although only *Importance* measure satisfied the threshold value for staff. Therefore, the expectation would be that most of the measures from Task 3 would meet the hypothesis. It is observed that this disparity was caused by the largely neutral¹⁰ rating assigned by the staff for these measures. In the system-level evaluation, the measures under *Perceived Usefulness* construct differentiated the two groups in both agreement percentages comparison and hypothesis testing. At an overall level, majority of the measures had agreement percentages above the threshold value for the students group, albeit the hypothesis was proved only for a few measures since the difference in the mean values were not large enough. Ideally, all the measures should satisfy the threshold percentage while not meet the hypothesis at the same time. This stage can be reached when the system manages to meet all the expectations of staff group.

¹⁰ The value '3' corresponds to Neutral in the Likert scale

Table 5.21

Consolidated results for the agreement percentages and hypothesis

Agreement Percentage > 75%		Hypothesis Met	
Students	Staff		
Task 1	Relevance, Good_Spread, Diversity, Recency, Usefulness, Expansion_Required, User_Satisfaction Seedbasket_Similarity,	Relevance, Usefulness	Relevance, Good_Mix, Novelty, Good_List
	Shared_Corelations, Relevance, Good_Spread,	Seedbasket_Similarit y, Relevance, Usefulness,	Seedbasket_Similarity, Shared_Corelations, Relevance,
	Diversity, Usefulness, Good_List,	Seedbasket_Usefulne ss	Good_Spread, Good_List
	Seedbasket_Usefulness Relevance,Usefulness,		
	Task 2	Importance, Good_List, Shortlisting_Feature	Importance
System	Task_Interconnectivity, Convenience (EUS*), Pleasant_Experience (PSE+), Awareness	Task_Interconnectivit y, Convenience (EUS)	Awareness (PSE), Accomplish_Tasks (PU), Performance_Improvabil ity (PU),

Agreement Percentage > 75%		Hypothesis Met
Students	Staff	
(PSE), Accomplish_Tasks (PU^), Performance_Improvari- lity (PU), Enhance_Effectiveness (PU), Ease_Job (PU), Work_Usefulness (PU)		Productivity_Improvari- ty (PU), Ease_Job (PU)

*Note: * - Effort to use the System, + - Perceived System Effectiveness, ^ - Perceived Usefulness*

Predictors for Agreeability on a Good list (Good_List)

The independent variables (evaluation measures) that had statistically significant predictive ability over *good_list* measure are listed in Table 5.22. *Usefulness* was the only measure present across all the tasks. This finding along with the previous findings, underlines the value of this measure in user evaluation studies conducted in the domain of scientific paper recommendations. Future user studies should include this evaluation measure since it has high prediction capacity across different tasks. The predictors for each task signify the nature of the task and the corresponding expectations. In Task 1, the measures *recency*, *novelty* and *serendipity* signify the exploratory nature of the task while the measure *seedbasket_similarity* signifies the directed search pattern of users in Task 2. In Task 3, the two measures *relevance* and *certainty* were the key predictors. Users, who felt that the recommendations can be

surely cited in their manuscript, preferred the recommendations list. In future studies with the Rec4LRw system, the generated models will be retested for validity purposes. It is expected that these measures will remain the key predictors along with the inclusion of new measures.

Table 5.22

Predictors identified in the regression models of the three tasks

Task	Independent Variables
Task 1	Recency, Novelty, Serendipity, Usefulness, User_Satisfaction
Task 2	Seedbasket_Similarity, Usefulness
Task 3	Relevance, Usefulness, Certainty

5.4.5.7 Inter-Task Analysis

In this set of analysis, the evaluation measures that were common across the three tasks are analyzed using the agreement percentages.

Transition in Participants' Perception from Task 1 to Task 2

In a realistic scenario, researchers use the seed papers from the reading list to find topically similar papers that help in formulating the exact research problem. The transition from exploratory search to directed searching happens in this second task. Correspondingly in this study, it was expected that the participants' evaluation of output quality measures will improve from the reading list task (Task 1) to the similar papers task (Task 2). Figures 5.14 and 5.15 illustrate the agreement percentages of students and staff respectively for four output quality measures and the *expansion_required* measure. T1 refers to reading list task and T2 refers to similar papers task in the figures.

For students group, there was improvement for two of the three output quality measures. The biggest improvement was observed for *good_list* (from 59.68% in Task 1 to 79.03% in Task 2). The only measure with drop in agreement percentage from Task 1 to Task 2 was *user_satisfaction* (from 75.81% in Task 1 to 70.97% in Task 2). More importantly, the drop for the measure *expansion_required* (from 82.26% in Task 1 to 74.35% in Task 2) indicates that participants' intent to get more papers, reduced in Task 2.

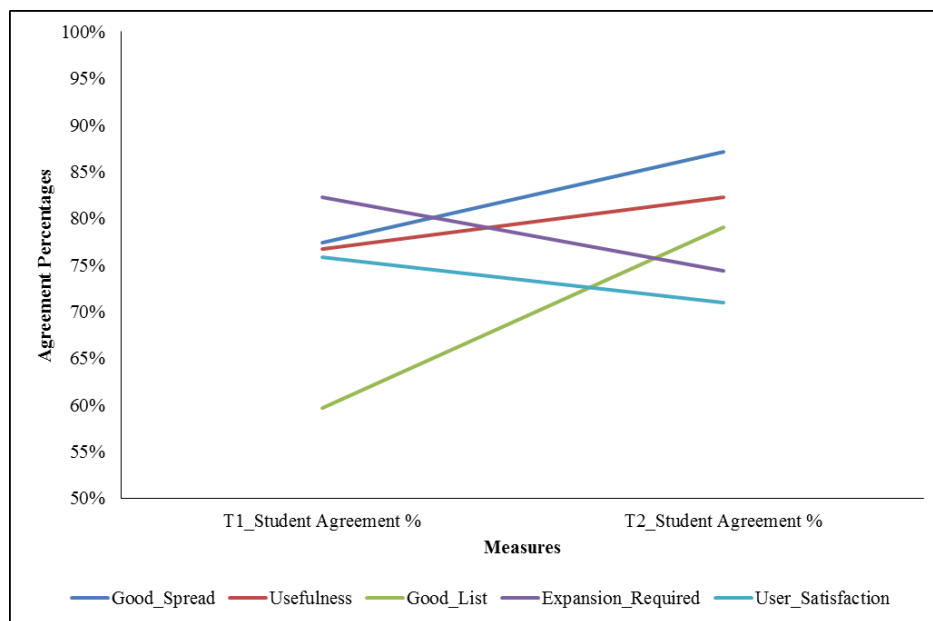


Figure 5.14. Comparison of five measures between the two LR tasks for students group

In the case of staff, there was considerable improvement for three output quality measures *good_spread* (from 64.29% in Task 1 to 73.77% in Task 2), *good_list* (from 52.86% in Task 1 to 65.57% in Task 2) and *user_satisfaction* (from 61.43% in Task 1 to 70.49% in Task 2). *Usefulness* (78.97% in Task 1 and 78.69% in Task 2) and *expansion_required* (71.43% in Task 1 and 72.35% in Task 2) were largely similar between the two tasks. The results for the staff group clearly indicated

improvement in finding relevant papers as participants zoom into the research sub-topic of their choice. The slight increase in *expansion_required* measure was because of the requirement of the participants in increasing the number of recommendations in the task.

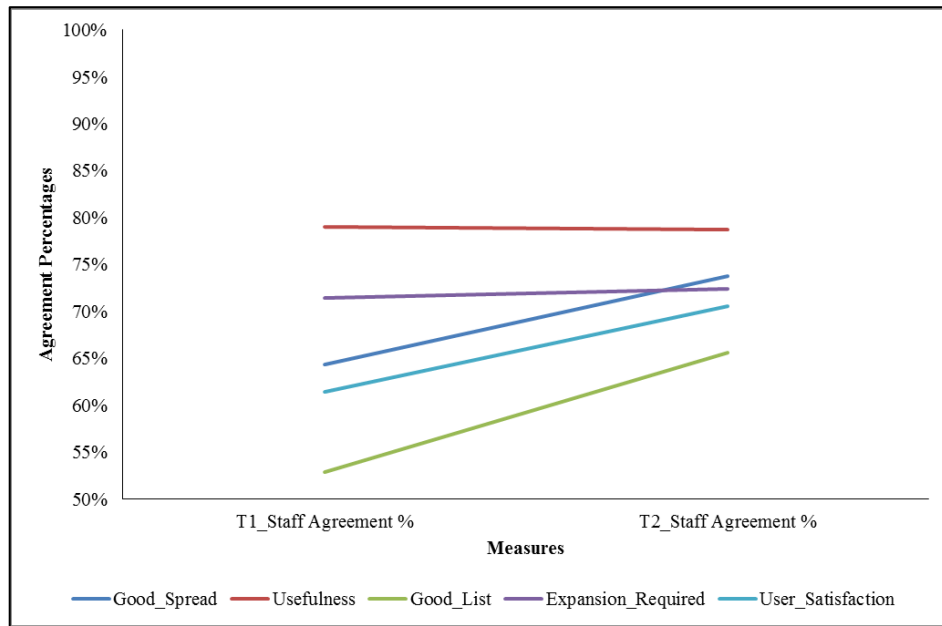


Figure 5.15. Comparison of five measures between the two LR tasks for staff group

Common Characteristics across the Three Tasks

In Figure 5.16, the agreement percentages of the three common evaluation measures are illustrated. For the *relevance* measure, the agreement was quite high in all the three tasks with Task 2 being the highest (90.32% for students, 83.61% for staff). Task 3 values were comparatively low as some of the participants were unsure of the applicability of the shortlisted papers to their article-type preference. In terms of the *usefulness* measure, the agreement was high for all cases with the exception of Task 3 staff group where the agreement was only 64.15%. There was a considerable difference of 17.85% between the two groups for this measure in Task 3. This drop

corresponds to the staff group's reluctance in citing the shortlisted papers, as indicated by the percentage on the *certainty* measure (61.02%).

Interestingly, the *usefulness* agreement in Task 1 for staff (76.72%) was found to be higher than students (78.97%). This finding validates the results for the measures *popularity* and *interdisciplinarity* for the same task (refer Figure 6.6). The inference here is that participants found the list to be useful due to the presence of popular and interdisciplinary papers. The measure *good_list* is an interesting case as the agreement for both groups in Task 1 (59.68% for students, 52.86% for staff) was considerably lower than Tasks 2 and 3. For the same Task 1, this measure's percentage was even lower than *good_mix* measure (70.97% for students, 58.77% for staff). One of the key reasons for this finding is the perceived low number of papers in the generated reading list. Most of the participants felt that the recommender paper's count was low for a reading list. Participants wanted a control feature in the UI for setting the number of papers. Among the three measures, *good_list* separated the groups with a substantial difference of more than 14% for Tasks 2 and 3, thereby clearly highlighting the higher expectations from the staff group. Staff concerns on issues such as providing control features in the user interface and improving recommendation quality of the tasks are to be addressed so that staff agreement shows improvement in the future.

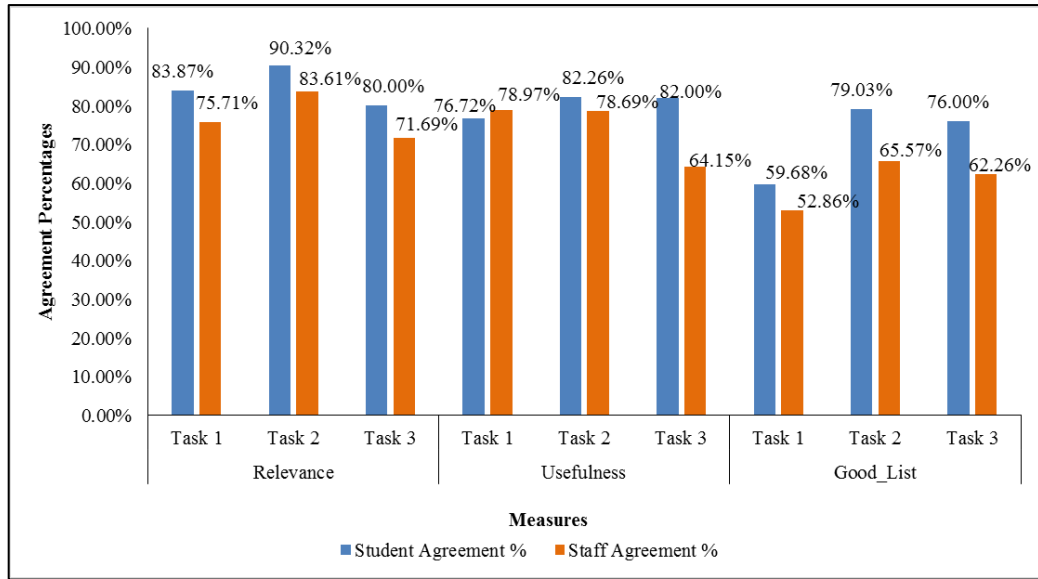


Figure 5.16. Agreement percentages of common task measures

Comparison of Pre-study and Post-study Participants' Opinion

In Figures 5.17-5.19, a clustered bar chart is illustrated for facilitating the comparison between the pre-study and post-study measures for the three tasks. The pre-study measure is *need_assistance* where participants indicated the frequency of needing external assistance while executing the respective task. The post-study measure is the measure *good_list*. In the X-axis, the *need_assistance* measure is depicted with the lowest and highest values as 1 and 5 respectively. The scale values are 1 (Never), 2 (Rarely), 3 (Sometimes), 4 (Very Often) and 5 (Always). The count is depicted in the Y-axis of the figures. The clustered bars at each frequency level refer to the participants ratings for the *good_list* measure. For the current analysis, only values 4 and 5 from the *good_list* measure will be considered since these values correspond to participant agreeability.

For Task 1 (refer Figure 5.17 and Table 5.23), participants who rated the frequency as 4 (Very Often) and 5 (Always), subsequently rated the list as a good list

at 61.29% (19 out of 31) and 78.57% (11 out of 14) respectively. Participants, who faced the frequency as 3 (Sometimes), subsequently rated the list as a good list only at 48.15% (26 out of 54). Therefore, the participants who faced the issue on a regular basis, found the recommendations to be good albeit the recommendations need to be improved for those participants who are seldom in the need for external assistance.

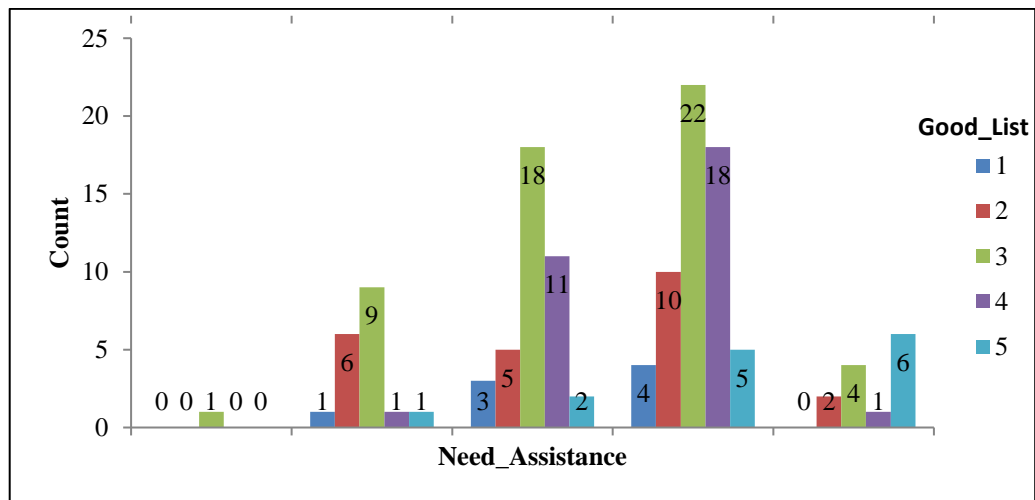


Figure 5.17. Pre-study and post-study comparison of Task 1

Table 5.23

Crosstab of pre-study and post-study variables of Task 1

Need_Assistance	Good_List					Total
	1	2	3	4	5	
1	0	1	3	4	0	8
2	0	6	5	10	2	23
3	1	9	18	22	4	54
4	0	1	11	18	1	31
5	0	1	2	5	6	14
Total	1	18	39	59	13	130

From the Figure 5.18 and Table 5.24 for Task 2, the results were slightly different from that of Task 1. The participants who rated the frequency as 3 (Sometimes), found the list to be good at 73.91% (34 out of 46). The participants who rated the frequency as 4 (Very Often) and 5 (Always), found the list to be good at 69.70% (23 out of 33) and 50% (3 out of 6) respectively. The overall results for this task are better than Task 1 since participants found the recommendations to be of good quality, regardless of whether they needed external assistance for this task.

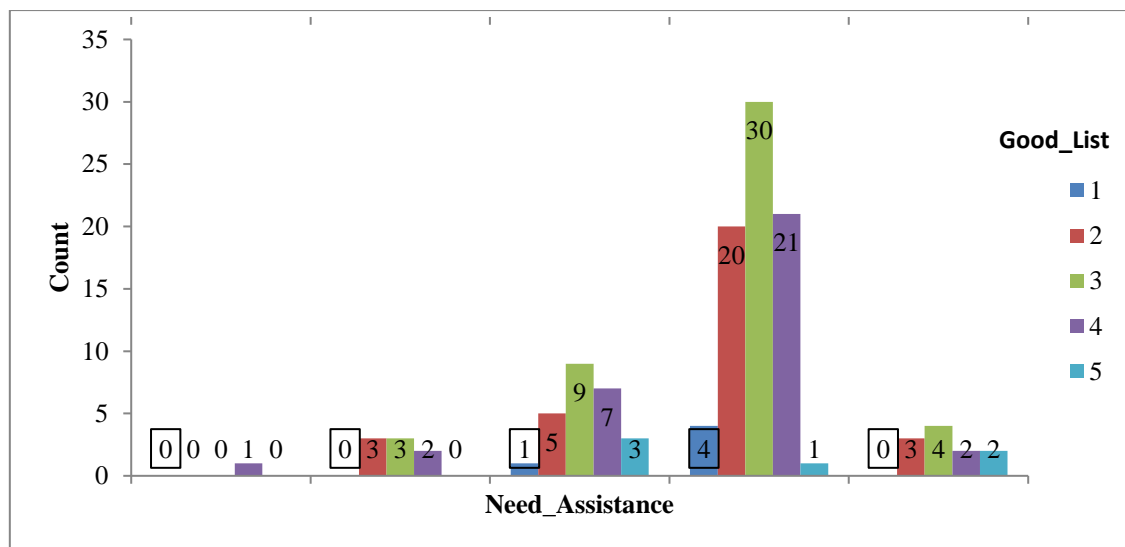


Figure 5.18. Pre-study and post-study comparison of Task 2

Table 5.24

Crosstab of pre-study and post-study variables of Task 2

Need_Assistance	Good_List					Total
	1	2	3	4	5	
1	0	0	1	4	0	5
2	0	3	5	20	3	31
3	0	3	9	30	4	46

4	1	2	7	21	2	33
5	0	0	3	1	2	6
Total	1	8	25	76	11	121

The results of Task 3 were the best among the three tasks for the pre-study and post-study opinion comparison (refer Figure 5.19 and Table 5.25). The participants who rated the frequency as 3 (Sometimes), 4 (Very Often) and 4 (Always), found the list to be good at 73.17% (30 out of 41), 76% (19 out of 25) and 60% (6 out of 10) respectively. These findings indicate a level of consistency across the different frequencies. This comparison analysis showed that a majority of the participants, who indicated the need for external assistance during the pre-screening survey, were benefitted by the recommendations.

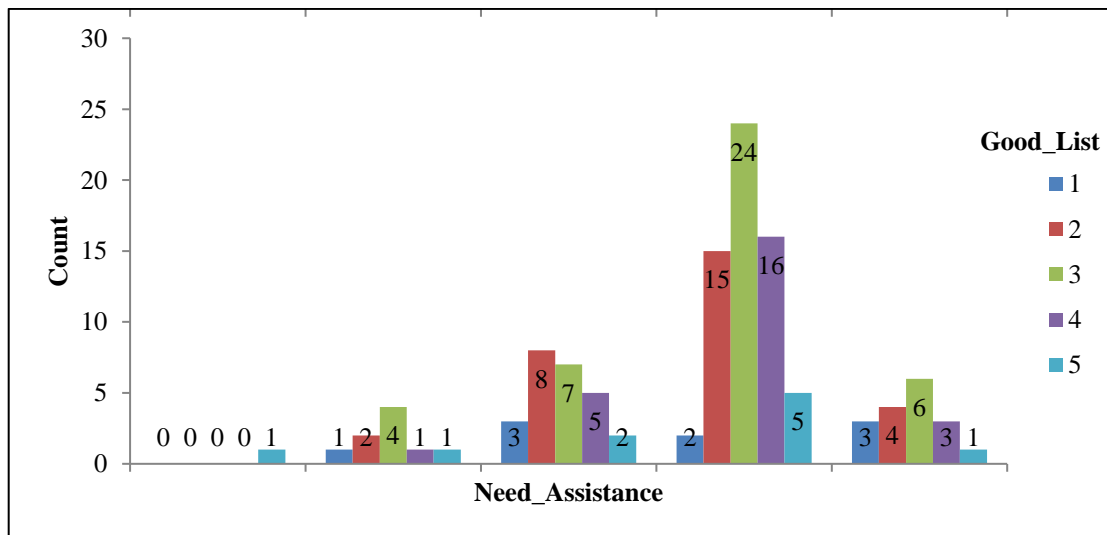


Figure 5.19. Pre-study and post-study comparison of Task 3

Table 5.25

Crosstab of pre-study and post-study variables of Task 3

Need_Assistance	Good_List					Total
	1	2	3	4	5	
1	0	1	3	2	3	9
2	0	2	8	15	4	29
3	0	4	7	24	6	41
4	0	1	5	16	3	25
5	1	1	2	5	1	10
Total	1	9	25	62	17	114

5.4.6 Qualitative Data Analysis

During the tasks and system evaluation, participants were requested to provide subjective feedback in the survey questionnaires. For each task, two questions were provided. The questions were (i) *From the displayed information, what features did you like the most?* and (ii) *Please provide your personal feedback about the execution of this task.* In the final questionnaire where the evaluation for both Task 3 and system was performed together, a single question was posted for eliciting feedback about the overall system. The question was - *Please provide feedback on the system. You can mention features that you liked in the system and also other features that could be added.*

The feedback data from these questions were coded using an inductive approach (Thomas, 2006) with the aim of identifying the central themes (concepts) in the participant responses. More specifically, the *descriptive coding* method was used

due to its relevance to this study. Saldana (2009) states “descriptive coding summarizes in a word or short phrase – most often as a noun – the basic topic of a passage of qualitative data”. For each participant comment, a primary code was mandatorily assigned and a secondary code was assigned if multiple themes were present in the comment. Two coders performed this exercise. The first coder identified the codes and these codes were then used by the second coder. Using the Cohen’s kappa statistic (κ) (Cohen, 1960), the inter-coder reliability was identified for the preferred and critical aspects of the three tasks while overall feedback codes were identified for the system level feedback comments. The κ values for the three tasks are listed in Table 5.26 while the κ value for the overall system feedback was 0.812. As per (Altman, 1990), κ value range of 0.61 to 0.80 can be considered to be good while the range from 0.81 to 1 can be considered to be very good in terms of the agreement strength. The κ value was consistently good for this study, thereby indicating sufficient agreement between the two coders. The full list of the resultant codes is provided in Appendix I. Based on the coded data, the two important aspects are presented – preferred aspects and critical aspects. The categories under preferred aspects are about the features which were appreciated or preferred by the participants. The categories under the critical aspects are about the features which needed improvement as per the participants’ opinions.

Table 5.26

Inter-coder reliability statistics

	Kappa Statistic (κ)	
	Preferred Aspects	Critical Aspects
Task 1	0.918	0.727

Kappa Statistic (κ)		
	Preferred Aspects	Critical Aspects
Task 2	0.930	0.758
Task 3	0.877	0.902

5.4.6.1 Preferred Aspects

In Table 5.27, the top five categories of the preferred aspects for the tasks are listed.

The percentages alongside the categories indicate the frequency of occurrence among the total set of comments for that task.

Table 5.27

Preferred aspects categories of the three tasks

Rank	Task 1 ($n=109$)	Task 2 ($n=100$)	Task 3 ($n=91$)
1	Information Cue Labels (41%)	Shared Co-citations & Co-references (28%)	Shortlisting Feature & Recommendation Quality (24%)
2	Rich Metadata (21%)	Recommendation Quality (27%)	Information Cue Labels (15%)
3	Diversity of Papers (13%)	Information Cue Labels (16%)	View Papers in Clusters (11%)
4	Recommendation Quality (9%)	Seed Basket (14%)	Rich Metadata (7%)
5	Recency of Papers (4%)	Rich Metadata (9%)	Ranking of Papers (3%)

Note: n indicates the total number of participants who provided feedback for the particular task.

In Task 1, the top two categories Information Cue Labels and Rich Metadata are related to the information display features in the Rec4LRW system. These two categories were present in the top five categories for all the tasks. Information cue labels (popular, high reach, survey/review and recent) helped the participants in quickly identifying the unique papers in the recommendation lists. The usefulness of these labels was indicated in all the three tasks by the participants, thereby making them applicable for implementation in other LR search tasks. The usage of cue labels could be extended beyond paper types. For instance, other types of labels such as interdisciplinary, high-impact journal/conference and altmetric score can be added to papers so that the uniqueness of papers in the recommendation list is conveyed visually.

“There was a categorization of papers as high reach, survey etc. This is very useful for beginners to understand which papers to read in the current context of this project.” –

[R01¹¹]

“The tags associated with papers wherever they are present allows for easy classification.”– [R58]

Under the *Rich Metadata* category, the metadata fields - abstract, references count and citations count were considered beneficial by the participants. It is to be noted that references count is displayed only in certain academic search systems. The presence of these basic metadata fields is mandatory for such systems. In the user evaluation of prior studies (Ekstrand et al., 2010; Jardine, 2014; Mcnee, 2006), this set of information was rarely displayed to the participants.

¹¹ RXX is a pseudonym pattern for the respondents in the evaluation study. R refers to respondent and XX refers to the number.

“The system provided the author specified keywords which let the users grasp the content of the paper in shorter time. The system has good indication of survey paper.”

– [R91]

“I am glad that the detailed references can also be obtained by this recommendation system.” – [R05]

The category *Diversity of Papers* corresponds to the requirement R4 of a reading list (refer Section 4.3.3.1 of Chapter 4). Participants were expecting to see a broad set of papers as a part of Task 1 recommendations since such papers would give them access to the subtopics in the research area.

“It lists papers with succinct details, any researcher would like to have such exhaustive list of papers spoon fed, though i understand this list was limited to ACM DL” – [R67]

“...provide interesting sub-topics that I'm not expecting to see”– [R04]

The category *Recommendation Quality* refers to the overall quality of the recommendations. Participants, who posted comments relating to this category, most probably agree to the task redesign intervention of the Rec4LRW system.

“I found that the papers were relevant to the topic which I selected which I think is important particularly because useful information gathering and organization is becoming extremely important these days for research.” – [R69]

“It also contains papers which are related to some Applications of using Sensor Networks which I was looking for.”– [R48]

The category *Recency of Papers* corresponds to the requirement R3 of a reading list (refer Section 4.3.3.1 of Chapter 4). Participants were expecting to see

recent papers as a part of Task 1 recommendations since such papers provide overview of the most recent research performed in the particular research area. It is to be noted that *recency* is one of the statistically significant predictors for the *good_list* measure in the multiple linear regression test results.

“The list of recent papers which are not easily shown in other paper search engines...”– [R30]

“The one which shows recent as well as high research area is something which I personally found very interesting as this also gives me an idea on which topic is my research area is highly researched these days.”– [R25]

In Task 2, the *Shared Co-citations and Co-references* category feature was perceived useful as it showed the relations between the recommended papers and the seed basket papers. This feature similar to the cue labels is novel and is not available in current search systems. The information in this feature could be further augmented with a visual citation network that places the seed basket papers and the recommended papers as nodes so that the relations are comprehended in a contextual manner. This feature is part of the informational display features intervention.

“The idea of showing shared citations and references is interesting. However, it would be interesting to see shared authors, shared conferences as it is highly likely that work by the same author might be very relevant/extending the seed works.”–[R01]

“Information about shared references and citations can help in establishing the relevance of these papers to the seed papers chosen.”–[R58]

The *Seed Basket* category corresponds to the comments where participants specifically stated that they liked the seed basket (SB) feature in Task 2. This feature is

part of the task interconnectivity intervention by which interaction is set between the Tasks 1 and 2.

“Seed basket helped to find more relevant papers to cover the topic. Shared co-citations is rather useful feature.”-[R29]

“The ability of giving a set of seed papers and asking the system to recommend related papers is a really nice feature. This could help in ensuring that we cite all the important/related papers in papers we write, and also to explore the field further.”-[R112]

In Task 3, *Shortlisting Feature & Recommendation Quality* was the top most coded category. This category corresponds to the task objective of shortlisting papers from the reading list and the resultant recommendations. Task 3 was a novel task for the participants since earlier approaches have concentrated less on tasks related to manuscript preparation/writing.

“I liked the concept of the system identifying possible papers for referencing for a specific event i.e. a conference, generic research paper etc..”-[R69]

“The process of generating a reading list is quite useful and I wish it was available in systems like Google Scholar. Good job!”-[R112]

The feature of viewing papers in the parent cluster of shortlisted papers (*View Papers in Clusters* category) was the third most preferred feature in Task 3 since participants found the shortlisting objective of the task and the information cue labels as the prominent favorites. Certain participants indicated that the feature helped them in discovering new papers in a serendipitous manner.

“The user can view many papers in the parent cluster in addition to the shortlisted papers. Thus the user need not spend much time on finding related papers.”-[R05]

“The view papers in the parent cluster function is very helpful to get a full picture of research field.”-[R64]

5.4.6.2 Critical Aspects

In Table 5.28, the top five categories of the critical aspects for the tasks are listed. The percentages alongside the categories indicate the frequency of occurrence among the total set of comments for that task.

Table 5.28

Critical Aspects Categories of the Three Tasks

Rank	Task 1 (n=109)	Task 2 (n=100)	Task 3 (n=91)
1	Broad topics not suitable (20%)	Quality can be improved (16%)	Rote selection of papers for task execution (16%)
2	Limited dataset (7%)	Limited dataset (12%)	Limited dataset (5%)
3	Quality can be improved (6%)	Recommendation algorithm could include more dimensions (7%)	Algorithm can be improved (5%)
4	Different algorithm required (5%)	Speed can be improved (7%)	Not sure of the usefulness (4%)
5	Free-text search required (4%)	Repeated recommendations from Task 1 (3%)	UI can be improved (3%)

Note: n indicates the total number of participants who provided feedback for the particular task

In Task 1, the category *Broad topics not suitable* refers to the issue where participants found the list of 43 input research topics to be broad for searching (e.g., machine learning, information retrieval). Since this study was conducted using an ACM DL corpus with a restricted set of papers, the free-text search option was not provided for the users.

“I may prefer a more specific research topic, because currently the recommendation list seems to cover a wide range in which some papers may be not interested.”-[R14]

“The research topics are very coarsely categorized; there is no way for me to get recommendations for more specific topics.”-[R33]

The category *Limited dataset* is attributed to the same issue as the previous category. Even though, participants were aware of the dataset restrictions, some of them made it a point to raise it as a concern during evaluation. Since the AKR technique relies on the author-specified keywords metadata of research papers, the corpus was restricted to the ACM DL dataset. The other publicly available datasets do not contain this metadata.

“Due to limitation of data sets (as only ACM papers) search result is not of decent quality. However, diversity and survey papers are useful.”-[R37]

“...Also the database is missing IEEE Xplore articles and from the best robotics journals and conferences which maintain high standards.”-[R02]

The category *Quality can be improved* corresponds to the participants comments where certain concerns were raised about the quality of the recommendations. Even though, the system managed to satisfy a majority of the students group, the staff group was not entirely in agreement with the recommendation logic in the techniques.

“A few of the recommendation are OK. Not covering many useful papers. Not covering some useful domains of application. Some of the papers are of peripheral importance.”-[R63]

“The recommendation is quite general (since machine learning is a very big topic). There's a need to improve the recommendation algorithm (e.g. base on the download count)...”-[R27]

Related to the previous category, *Different algorithm required* category is an interesting case since participants suggested alternative algorithms that could be utilized for the recommendations. Such comments represent the different information paths used by researchers in finding papers for their LR.

“Take a high impact paper (based on citation and may be exact keyword matching), then go through its own references to understand more about the research conducted. This is because, a good work generally cites other prominent works in the field.”-[R02]

“It would be much better if you follow the embedded system classifications, suggested by IEEE/ACM and pick up a relevant paper from each class. The other option is to simply browse the papers from top journals/conferences, which have an embedded systems track and report those. In short, there is not enough domain-specialization yet in this tool to make it useful...”-[R03]

The category *Free-text search required* refers to a known restriction with the user evaluation study. Since the participants are used to free-text search in contemporary search engines, the absence of this feature was an issue for most participants. The next release of the Rec4LRW system should have this feature so that users could freely express their information needs through search keywords.

“The execution of this task is simple and the displayed results are clear and obvious.

However, if there is a sub-function to allow more specified searching, such as key words, it will be better.”-[R10]

“However, as much as I like the system, I still feel that there should be a provision for the user to enter specific search queries as opposed to merely searching by topics.

This would further advance the existing system...”-[R69]

In Task 2, another aspect of algorithm improvement was raised with the category *Recommendation algorithm could include more dimensions*. Some of these comments relate to the need for increased user control in the interface along with the more input options so that users are able to use the system to the maximum extent.

“I think you need to collect more information from the user, such as keywords. Also you need to let the user remove the unrelated papers from the list, so that you exclude the unwanted topics.”-[R57]

“Can you short list them based on the seed papers ? This is because seed papers might be from different areas, so if you can sort it based on its areas, it will be pretty helpful.”-[R55]

For the category *Speed can be improved*, the retrieval speed was an issue for some participants, particularly in Task 2. Since the IDSP technique in Task 2 finalizes the recommendations from three different paper discovery modules, the overall retrieval time is comparatively higher. Users enjoy fast retrieval speeds in systems such as Google Scholar and ScienceDirect, therefore they would expect a new system to have similar capability.

“The time spent on generating the recommendation list was a little bit longer than expected”-[R14]

“Too slow, it takes more than half a minute to generate the recommendations which is unbearable.”-[R47]

In Task 3, *Rote selection of papers for task execution* category corresponds to the top critical issue. Due to the nature of the study design, participants were requested to include a minimum of thirty papers for the final reading list which is the main input for the task. The selection of papers solely for the purpose of executing the task was perceived as a concern by some users. However, this issue does not provide scope for mitigation unless the study is conducted in real-world settings as a longitudinal study.

“Had to choose irrelevant papers in my reading list to reach the 30 mark level”-[R12]

“This was a hectic and time consuming task, it takes long time, don't know why the selection of 30 was so important?”-[R85]

The category *Not sure of the usefulness* refers to the comments from participants who were not sure of the usefulness of Task 3. In the user guide provided to the participants, an explanation on the usability of the task was provided. Nevertheless, the task was not construed as useful by four participants.

“I'm not sure whether there is really that important to shortlist papers based on article-type preference”-[R04]

“I'm afraid I am not too clear what the purpose of this stage was. The hypothetical task of writing a research paper on "social media" is just too vague to pin down in a survey such as this.”-[R19]

5.4.6.3 System-level Categories

The top five categories coded for the system feedback, are listed in Table 5.29. These categories are a combination of both preferred and critical aspects of the overall system. The complete list of categories is provided in Appendix I.

Table 5.29

System-level categories

Rank	System-level Evaluation Aspects ($n=91$)
1	Satisfied with system (16%)
2	Free-text search required (13%)
3	Broad topics not suitable (8%)
4	Limited dataset (6%)
5	Seed basket is good (5%)

Note: n indicates the total number of participants who provided feedback for the particular task

The top category *Satisfied with system* indicates a majority of the respondents had no issues with the system and found it to be quite useful for the purpose it has been built. The comments from the participants indicated that both the recommendations and the user-interface of the system (mainly facilitated by seed basket and reading list) made the system functional in a convenient way.

“The concept is very good.. I like the way the system replicates the actual process of conducting research from early literature search (seed search) to a more comprehensive list(reading list). I like the fact that the system is designed to cater to the user at each stage.. I feel the system can be more robust if the user can also enter

specific search queries in addition to broad research topics so as to further refine the search.”-[R69]

“The initial pain of literature survey which is time-consuming and tedious is reduced to a great extent!!!”-[R92]

The comments for the next three categories *Free-text search required*, *Broad topics not suitable* and *Limited dataset* were repeated by the participants from the tasks evaluation. This repetition indicates that these issues affected their evaluation to be a large extent. Even though, these issues were expected to be raised by the participants as they are due to the study design, the level of impact was ascertained only at the end of the study. A major implication for future studies is that free-text search should be provided to the participants during evaluation since a list of topics would be deemed insufficient. The fifth category *Seed basket is good* is an indication of the popularity of this task interconnectivity mechanism. This category was one of the preferred aspect categories from Task 2.

5.5 Discussion

This section presents a discussion of the analyses from the user evaluation study which examined whether the recommendation techniques and the Rec4LRW system met user expectations. The discussion is split into three sub-sections for the three design interventions which were employed for the design and development of the Rec4LRW system.

5.5.1 Influence of Task Redesign Intervention

The task redesign intervention basically entails the recommendation techniques of the three tasks. One of the expectations before the study was that the redesigned aspects of

the tasks i.e. the novel characteristic(s) corresponding to the requirements of the tasks, would influence the participants more than the generic characteristics. Task 1 results indicate that *good_spread*, *diversity* and *recency* measures were well received by the student participants. The importance of diversifying recommendations has been raised in previous SPRS studies (Küçüktunç et al., 2015) and in Task 1's context, a range of papers is deemed to be beneficial for researchers at the start of LR. In addition, recent papers are an interesting case since previous SPRS studies (Bae et al., 2014; Ekstrand et al., 2010) have focused on seminal papers specifically. It is to be argued that recent papers are as important as seminal papers since researchers would want to know about the most recent research performed in a particular research area. The scientific information seeking model (Ellis & Haugan, 1997), Ellis alludes to the search for papers that help in maintaining awareness of research topics during the *Browsing* stage. Hence, recent paper recommendations are important. On the flip side, three issues hampered the user experience in this task - the lack of free text search, limited coverage of the dataset (restricted to ACM DL) and fixed recommendations count. While the former two issues were implications of the study design, the participants' observation on the limited number of recommendations was an interesting case since most of the previous studies offered a maximum of 20 recommendations per task. It can be argued that users should control the number of recommendations, although there is the caveat of recommending low quality papers if the number is set very high.

Task 2 results were the most promising since there was very less differences between the two groups. Even though, participants indicated that this task took approximately one minute for generating the recommendations which is a high wait time, they were not perturbed by the recommendation quality. Similar to Task 1, the measures *good_spread* and *diversity* had high agreement percentages, thereby

validating the ability of the paper discovery methods in the IDSP technique. More importantly, the task was able to retrieve recommendations similar to most of the papers in the seed basket. Interestingly, many participants felt that the quality can be still improved by incorporating alternative approaches and more input dimensions. It has been already highlighted in Chapter 2 that this task had the most number of previous studies among all SPRS studies. This is due to the observation that this task is a very ambiguous task as similar papers could be found on multiple grounds. Perhaps, sub-tasks could be designed for this task where each sub-task is assigned a fixed goal (e.g. find the most similar papers from the citation network of a particular paper in a seed basket of multiple papers). However, the ambiguous nature of this task from a user viewpoint might warrant the inclusion of “Explanations” alongside the recommendations. ‘Explanations in RS’ is an area of research where the objective is to conceptualize methods for providing explanations to end-user on why particular recommendations are made (Tintarev & Masthoff, 2007). This intervention can improve the user experience of researchers particularly for this task. Notwithstanding, the incorporation of semantic textual similarity methods (Han, Kashyap, Finin, Mayfield, & Weese, 2012) in the future version of the IDSP technique provides a possibility to further improve the quality of recommendations.

Task 3 results were promising in lieu of its novelty in SPRS studies. Even though, participants’ responses indicated the shortlisted papers were important papers, they were still unsure about citing these papers in their manuscripts. This observation perhaps highlights the limitation of the task since the citing behavior of researchers is very much based on their personal context. Erikson & Erlandson (2014) had identified that researchers mostly cite papers for supporting their claims and for evidential purposes. Therefore, citation context based recommendations (He et al., 2011) would

be more suitable for researchers. However, this task was deemed to be useful in a situation where the reading list of papers collected during the LR is a big list as finding unique and important papers from this list would be a manually complex task. The positive responses from the participants for the shortlisting feature is an encouraging sign as they many participants explicitly stated the usefulness of this task if incorporated in current academic search systems. Hence, this task can be integrated with the citation context recommendations task for further improving the overall utility of SPRS systems for researchers. Overall, the influence of the task redesign intervention on the Rec4LRW system was positively highlighted by the participants albeit with less intensity by expert researchers (staff).

5.5.2 Influence of Task Interconnectivity Intervention

The task interconnectivity mechanisms in the Rec4LRW system basically ensure that there is connectivity established across the three tasks through paper collections. The seed basket (SB) and the reading list (RL) were incorporated in the Rec4LRW system for this purpose. The study results indicate an overwhelming support for these features in terms of usability and effectiveness. The results mirror the success of the recently developed *SearchAssist* system (Huurdean et al., 2016) where the impact of search user interface (SUI) features were evaluated for different sub-tasks. It has been highlighted that current academic search systems and digital libraries do not provide the necessary interface for managing different types of LR search tasks (Du & Evans, 2011b). Therefore, the positive feedback for these two features was on expected lines. In the recent years, academic search engines such as Google Scholar and Semantic Scholar have introduced personal paper collection features for users to save papers. The saved papers can be then used to find related papers. Thus, the importance of paper collections is being recognized by the research community. Based on the current

research, future systems which are supposed to handle multiple LR search tasks would be best served if SB and RL are incorporated as mandatory features.

5.5.3 Influence of Information Display Features Intervention

As highlighted earlier in Section 5.4.3.1 of Chapter 5, certain novel informational display features in the Rec4LRW system were included in the system for highlighting the redesigned task aspects. Similar to the positive responses for the previous intervention, this intervention was also well received by the participants. The effectiveness of these features was ascertained only through the subjective feedback of the participants. The information cue labels were particularly well endorsed by the participants since these labels expedited the relevance judgment decision for them. The impact of labels can be attributed to Zipf's principle of least effort (Zipf, 1949) since the cognitive load on the participants was minimal. The other two most appreciated features: shared co-relations and 'view papers in parent cluster' were effective in their respective purposes. The former showed the relations between the recommended papers and the seed basket papers (in Task 2) while the latter showed why a particular paper has been shortlisted (in Task 3). These display features served as a cognitive bridge between the recommendations and the participants while ultimately highlighting the usefulness of informational features in user-interfaces (Wilson, 2011). The implication for SPRS studies is to provide importance to UI display features during user evaluation since it is equally vital to properly highlight the recommended resources to the users since users aren't expected to expend much effort.

5.6 Summary

A total of seven evaluation goals were developed in Study II, including one hypothesis. Six of these goals helped in investigating the influence of the threefold

intervention framework of the Rec4LRW system on researchers. As a part of the pre-study, the offline evaluation experiment evaluated the four requirements of the reading list with seven techniques. Four variants of the AKR technique were benchmarked against three baseline techniques. The HITS-enhanced variant of the AKR technique was identified to be the best performing technique as it satisfied most of the requirements.

Through the user evaluation study conducted with 132 researchers, it was convincingly established that students preferred the task recommendations and the overall system. This observation was supported by the higher agreement percentages for most of the evaluation measures. On the other hand, staff participants found the system to be useful albeit less effective. However, the hypothesis was not met for a majority of the evaluation measures since the difference in the ratings of the two groups was not substantial, thereby proving favorable for the system. Task 1's results were affected by two issues namely limited dataset and lack of a free-text search option. Task 2 was the best rated task among the three tasks since it is a task which is manually complex in terms of execution. Task 3 being a novel task in SPRS studies, was well appreciated by the participants, thereby validating the effort put in this research to introduce a recommendation task for the manuscript preparatory stage of the research lifecycle. The measures which were identified as the most significant predictors for the *good_list* measure in the three tasks were found to be representative of the nature of the corresponding tasks. However, the validity of these predictive measures needs to be tested in future studies.

From the subjective feedback of the participants, the effectiveness of the threefold intervention framework was specifically ascertained by positive comments on the usefulness of seed basket, information cue labels and the shortlisting feature of Task 3. Participants' feedback comments were found to be both critical and insightful

as their comments were based on a thorough usage of the system through the evaluation study. RQ3 was successfully addressed through the offline evaluation of Study II while RQ4 was partially addressed since the system and the task recommendations successfully met the expectations of research students. Future work needs to concentrate on satisfying staff participants so that the system becomes suitable for all demographics.

CHAPTER SIX - CONCLUSION

6.1 Introduction

This chapter concludes the dissertation with a discussion on the contributions and implications of this current work for research and practice, its limitations, and recommendations for future studies. The chapter is divided into four sections. The first and second sections present a summary of the completed research along with the contributions. The implications for research and SPRS implementation are discussed in the third section. The limitations of this research are subsequently presented along with the potential future works in the final section.

6.2 Review of PhD Research

Literature review and dissemination of research are two of the key activities in the scholarly research lifecycle (Nicholas & Rowlands, 2011). The former helps in synthesizing existing literature and finding research gaps while the latter helps in establishing the contributions of research. The skills and procedural knowledge required for the execution of these activities are learned through acquisition of the theoretical knowledge and learning by experience. Studies have shown that novice researchers such as undergraduate and graduate students struggle in most phases of LR (Du & Evans, 2011b; Spezi, 2016). As a result, they are found to be dependent on *people-oriented* and *technology-oriented* interventions. *People-oriented* interventions are the classical interventions where librarians, expert researchers and PhD supervisors assist beginners in completing the tasks. Although these interventions are the most ideal, they are constrained by resource and temporal availability. *Technology-oriented* interventions bypass these two constraints although they are affected by other

constraints. Specialized IR and RS are examples of *technology-oriented* interventions that provide task-based recommendations to users based on algorithms which simulate the user search process. The efficacy of these algorithms is constrained by the operationalization quality of the corresponding task's heuristics. With reduction in computer peripheral costs and improvement in algorithms, these interventions have become one of the preferred options for researchers, regardless of their experience level. However, the importance of experts in providing tailored guidance to novices cannot be understated and should never be undermined.

With the intent of contributing towards *technology-oriented* interventions, this study began with the general question “*How can the most relevant research papers be delivered to researcher's literature review requirements?*”. The intent was to build a new scientific paper recommender system (SPRS) based assistive system. Studies from the related areas of scientific information seeking, scientific paper information retrieval (IR) and recommender systems (RS) were surveyed to get an understanding of the existing research and the possible opportunity areas. The survey of SPRS studies revealed a few issues at the structural and algorithmic level. Among the issues, there were certain issues which seemed important. The lack of connectivity between recommendation tasks and lack of intermediate set of features in the recommendation techniques of different LR search tasks was apparent. Thirdly, the usage of article-type for recommending papers during the manuscript preparatory stage, offered scope for utilizing SPRS in order to help researchers.

For addressing these issues, this research sought to build an assistive system for aiding researchers in acquiring papers for the literature review and manuscript preparatory tasks. The entire research was guided by two research objectives to achieve the aforementioned research goal. The first objective RO1 was to identify an

appropriate method to map the identified LR and MP tasks to relevant IR/RS algorithms. The second research objective RO2 was to evaluate whether the performance of the proposed recommendation techniques for the tasks and the overall system were at the expected level.

The research work was split into three phases. As part of Study I, a pre-design survey was conducted as a precursor for the prototype system development and evaluation. The central topic of the survey was set as ‘inadequate and omitted citations (IOC) in research manuscript’. 207 participants were surveyed about the frequency of occurrences of the different instances of IOC in journal and conference manuscripts, along with the reasons and effects of IOC. Popularity of academic information sources such as academic databases, search engines and stand-alone SPRS were also ascertained through the survey.

Based on the results of Study I, three tasks were selected for the prototype assistive system. Google Scholar’s UI design was adopted for the prototype since a separate study was not conducted for identifying user-interface (UI) components. The prototype was named Rec4LRW, an acronym for Recommender System for Literature Review and Writing. The three tasks selected for this system were (i) building an initial reading list of research papers, (ii) finding similar papers based on a set of papers, and (iii) shortlisting papers from the final reading list for inclusion in manuscript based on article-type preference of the researcher. In order to base the recommendation techniques on an intermediate set of features, seven features were identified. These features represent the characteristics of the bibliography and its relationship with the parent scientific paper. Out of the seven, three features *topical and peripheral coverage (TPC)*, *specificity* and *grey literature percentage* were novel features.

A task redesign exercise was performed for the three identified tasks for addressing the limitations of prior SPRS studies. The eventual design of the recommendation techniques of the three tasks was grounded on three different notions. The first task of building initial reading list was perceived as a *ranking problem* where the main objective was to rank a bigger list of papers based on a suitable pre-computed metric/rank. The second task of finding similar papers was perceived as an *extended paper discovery problem* where the main objective was to put together a hybrid technique that combined both textual and non-textual techniques for multiple input papers. The third task of shortlisting papers was perceived as a *cluster detection problem* where the intent was to identify clusters within the citation network so that the top papers could be shortlisted accordingly. Hence in effect, *information filtering* based approach was largely followed for the design of the recommendation techniques for the tasks. This approach is in contrast to the *machine learning* based approaches followed in few of the prior studies, thereby consciously circumventing the overhead time in training predictive models on a periodic basis. In the prototype system, two other novelties were introduced. Since there were three tasks, two paper collection features called as the *seed basket* and *reading list* were incorporated to enforce interconnectivity within the three tasks. Certain new informational display features were included as part of the UI for highlighting the unique aspects of the recommended papers. Thus, the three interventions - task redesign, task interconnectivity and informational display features were used in the design and development of the Rec4LRW system, thereby addressing the first research objective.

As part of Study II, two types of evaluations were conducted with the recommendation tasks and the Rec4LRW system – offline and user evaluation. The first task's recommendation technique was subjected to an offline evaluation

experiment where four variations of the proposed AKR technique were evaluated against three baseline approaches from earlier studies. The best performing technique was identified based on the ability to retrieve the expected paper-types from the corpus. The AKR technique provided the best results in the experiment conducted with papers retrieved for the top 186 keywords from the ACM DL corpus. A user evaluation study was conducted with the Rec4LRW system with 119 participants from 15 countries, predominantly from Asian universities. All the shortlisted participants had published at least one research paper; therefore the sample was very much representative for this study. The participants were split into two user groups – students and staff group. They were provided with a user guide and allowed to execute the study from any location since the Rec4LRW system was made accessible through the internet. The participants were expected to complete all the three tasks and answer three separate questionnaires. It was hypothesized before the start of the study that students would prefer the task recommendations and the overall system than staff. The evaluation data was extensively analyzed using agreement percentages and statistical tests for hypothesis testing, correlations and regression with ‘agreeability on a good list’ as the dependent variable, along with a pre-study and post-study variables comparison analysis.

The study results indicated that the task redesign intervention was a success at least for the student participants as there was high agreement on most of the relevant evaluation measures. The first task was preferred for the diverse and recent set of papers while the second task was preferred for its ability in finding a good spread of papers that were topically similar to the input seed basket papers. In the case of the third task, although participants agreed that the feature of shortlisting papers was a worthwhile idea, they were slightly hesitant on the citation aspect of these papers. The

other two interventions task interconnectivity and novel informational display features were unanimously favored by the participants since these features helped in a better user experience and faster relevance judgment decisions respectively. Overall, the task recommendations and the system design satisfied graduate research students to a large extent while there is definite scope for improvement so that staff participants' expectations are completely met in future versions of the Rec4LRW system.

In light of the findings of Study I and II, four journal papers, one technical report and five conference papers have been either accepted or published to date while one journal paper is currently under review. These are listed in Appendix J.

6.3 Contributions

This research mainly contributes to the area of scientific paper recommender systems (SPRS). Earlier dissertation studies on SPRS (Beel, 2015; Jardine, 2014; Mcnee, 2006; West, 2010) have provided solutions to different aspects of SPRS. In these studies, the opinions of researchers have not been collected before the design of the system and the techniques. The current research considers the experiences of researchers by collecting data on the topic of "Inadequate and Omitted Citations (IOC)". Therefore, the selection of tasks and a portion of the prototype (assistive system) design are based on researchers' perceptions. The dual perspective (manuscript reviewers and authors) based data collection approach used in Study I is an approach that can be used in future citation behavior studies. This approach facilitates data validation. From Study I, the criticality of the different reasons that lead to IOC, usage and awareness levels of traditional academic information sources and SPRS were identified. These findings should be useful for SPRS studies in technique conceptualization and raise a case for creating more awareness of specialized IR and RS systems.

Out of the seven base features proposed for usage in the recommendation techniques, *Topical and Peripheral Coverage (TPC)* and *Specificity* are novel features that can be used in algorithm design and paper ranking in future studies. *TPC* value of a paper indicates the ability of the paper in covering the papers published for the author-specified keywords based on the paper's citation network. *TPC* is similar to Eigenfactor (West et al., 2010) and Common Citation Inverse Document Frequency (CCIDF) (Lawrence et al., 1999) metrics due to its applicability in academic digital libraries. *Specificity* is a feature that can be used for measuring the non-textual similarity between paper and its citation, based on co-occurrences of the metadata fields - author-specific keywords and categories¹².

The recommendation techniques of the three tasks in the system are the most important contributions of this research. For Task 1 which is the reading list generation task, the AKR (Author-specified Keywords based Retrieval) technique has been proposed. The AKR technique can be utilized for providing popular, recent, survey and a diverse set of papers as a part of the initial reading list in literature review. For Task 2 which is about finding similar papers based on a set of input papers, the IDSP (Integrated Discovery of Similar Papers) technique has been proposed. The IDSP technique simulates manual paper discovery methods by combining citation chaining, textual and non-textual paper filtering mechanisms. Unlike prior techniques (Huynh et al., 2012; Liang et al., 2011) which are best suited for finding similar papers for a single paper, the IDSP technique can recommend similar papers for a bigger set of input papers. For Task 3 which is the novel task for shortlisting papers from the final reading list, the CNS (Citation Network based Shortlisting) technique has been proposed. The CNS technique makes use of

¹² This field is applicable only for certain digital libraries such as ACM Digital Library

community detection algorithm in shortlisting papers. This lightweight technique can be easily incorporated into reference management software such as Mendeley where researchers store research papers for their studies. Based on the Design Science Research (DSR) Knowledge Contribution Framework (Gregor & Hevner, 2013), the contributions to Tasks 1 and 1 can be classified as *Improvements* knowledge type¹³ while the contribution to Task 3 can be classified as *Invention* knowledge type¹³.

The threefold intervention framework proposed through this research provides a holistic approach to task-based assistive system design and development. The three interventions in this framework are task redesign, informational display features and task interconnectivity. In the Rec4LRW system, this framework was used to complement the proposed recommendation techniques of the three tasks with UI informational display features. Future studies can use the framework for systematically addressing multiple recommendation/retrieval tasks within an information system since it is a simple and easier exercise when compared with earlier approaches such as the 5S model for digital library development (Gonçalves, Fox, Watson, & Kipp, 2004).

From the user evaluation study of the recommendation techniques (Study II), the measures having predictive capacity over the main output quality measure “Good List” were identified. This is the first research study that has identified such predictive measures during user evaluation for recommendation tasks. The paper-types (e.g., recent, popular) corresponding to these measures could be given higher priorities in future SPRS studies. Furthermore, the measures that had high correlations with the output quality measures were also identified for the three tasks. These findings should

¹³ As per (Gregor & Hevner, 2013), *Improvement* knowledge type is developing new knowledge/solutions for known problems and *Invention* knowledge type is inventing new knowledge/solutions for new problems

help in earmarking the important characteristics for future SPRS from a user expectations viewpoint.

6.4 Implications

This study has yielded several contributions to research and practice in the fields of scientific paper recommender systems, as detailed in the following sub-sections.

6.4.1 Implications for Research

The findings of this research have theoretical implications. Study I is one of the first studies to look at the topic of inadequate and omitted citations (IOC), using a quantitative research method in Study I. Earlier studies (MacRoberts & MacRoberts, 1988, 2010) have predominantly used qualitative methods for this topic. The use of qualitative research method in previous studies was justified since the analyses were limited to documents. On the other hand, survey-based data collection in Study I was justified since user perceptions were to be collected for a fixed set of sub-topics under IOC. From Study I, it was identified that seminal and topically-similar papers were the most critical instances of IOC often missed by authors, across disciplines. This observation validates the existence of multiple RS studies conducted for these two instances and it stresses the need for digital libraries to provide alternative ranking options for users. Among the reasons for IOC, the lack of research experience in a particular area was perceived to be a major reason. On the contrary, the ‘lack of overall research experience’ is a recurrent theme from earlier information behavior studies. Future research can be conducted for delineating the differences between these two experience gaps. The awareness of specialized recommender systems for scientific papers was found to be very low among researchers. Hence, there is need to re-examine the visibility level of these systems along with the corresponding publicity

measures and the steps that are to be taken for creating more visibility as researchers keep reverting to the traditional information sources.

The task redesign exercise posited the importance of identifying the requirements of the tasks before designing the corresponding recommendation techniques. Earlier SPRS dissertation studies (Jardine, 2014; Mcnee, 2006) had not taken this aspect into account. The effectiveness of this exercise was ascertained from the Study II results where the novel characteristics of the tasks were highlighted by high agreement percentages for the corresponding evaluation measures. Hence, future SPRS studies should consider utilizing the business-process approach of breaking up a task into input, processing and output components (Sethi & King, 1998) to facilitate comparison with prior studies.

From Study II, regression results from the three tasks helped in identifying the predictors for the important variable – ‘Agreeability on a good list’. For Task 1, the five measures *recency*, *novelty*, *serendipity*, *usefulness* and *user_satisfaction* were found to be strong predictors. The absence of the *popularity* measure as one of the predictors for this task is a topic worth exploring since previous studies (Bae et al., 2014; Ekstrand et al., 2010) for this task have assumed that popular/seminal papers are the main paper-type expected of a reading list. A survey-based study could be conducted to ascertain the priority of the each paper-type for this task. This approach can be extended to the other two tasks as well. While for Task 2, the predictors were *seedbasket_similarity* and *usefulness*. The novel Task 3 had one task-specific and two generic predictors – *relevance*, *usefulness* and *certainty*. Even though, these relationships were statistically significant with an adequate sample size, these findings are to be validated in subsequent studies. Nevertheless, these measures can be given additional weightage while designing future SPRS-based assistive systems.

In addition, this research found that the information cue labels were perceived by participants of Study II as the most important and useful feature of the Rec4LRW system. Earlier SPRS studies have more or less ignored such display features as part of system design. In Rec4LRW system, the paper-types were depicted through the cue labels. However, further research needs to be conducted on the possible display features that can be incorporated into the UI. The inclusion of such features is to be seen in the same vein as the incorporation of “explanations” in RS (Tintarev & Masthoff, 2007). Explanations in RS help in informing the user about the reasoning logic behind the recommendations.

Although, the 22 evaluation measures identified for the three tasks were comprehensive in terms of evaluating different aspects of the recommendation lists, these measures operate at the list level and not at the item level. Therefore, the evaluation was carried out only at an aggregated level. It is to be seen whether an item-level evaluation approach would provide different results in the case of SPRS studies. In an earlier SPRS study (Jardine, 2014), a binary rating based item-level evaluation approach was employed, however list level evaluation was not carried out. The click-through rate (CTR) of recommended papers could also be used as a proxy measure for evaluating item-level recommendations. CTR was used in evaluating the Docear system (Beel, 2015).

6.4.2 Implications for Practice

The novel threefold interventional framework comprising of task redesign, informational display features and task interconnectivity is a framework that can be adopted in the design and development of future task-based assistive systems. The

three interventions complement each other and help in providing a more holistic user experience considering algorithmic, UI and task aspects.

The recommendation techniques from the three tasks are designed for use at digital libraries where the required metadata fields are indexed. Task 1 and 2 are suitable for direct implementation as these are important tasks that are executed by researchers during LR. Since the values for the features *TPC*, *textual similarity* and *specificity* are pre-computed, the recommendation process will be faster with low overhead when implemented in digital libraries. During the user evaluation study, participants found Task 2 to be the most useful among the three tasks since the execution of the task with the current systems is a difficult proposition. The cue for digital libraries is to extend the “Related Papers” feature for accepting multiple papers instead of a single paper. This suggestion is applicable for systems such as Scopus and Google Scholar.

The CNS (Citation Network based Shortlisting) technique from Task 3 can be utilized as a *plug and play* algorithm particularly in reference management systems such as Mendeley and Docear (Beel, Langer, Genzmehr, & Nürnberger, 2013). From the usage context, the CNS technique can be integrated with the global recommendations from Claper (Wang et al., 2010) so that subsequent citation contexts (He et al., 2011) recommendations could be made.

6.5 Limitations

Although this research yielded new and interesting insights, there are limitations that could be addressed in future research. From a methodological viewpoint, this research started with a survey on the topic of IOC in research manuscripts and other related sub-topics. The prototype system was designed and developed based on few findings

from the survey. However, the redesign of the tasks was not based on empirical data; instead it was based on the identified shortcomings. Similar to the research method used in Ellis scientific information seeking studies (Ellis et al., 1993; Ellis & Haugan, 1997), data from hard and soft science disciplines can be collected for establishing the heuristics behind the three tasks. It is postulated that there will be structural similarities across the disciplines.

In terms of algorithmic limitations and recommended improvement, the AKR (Author-specified Keywords based Retrieval) technique from Task 1 does not include any explicit preference to the four paper-type requirements. If popular papers count is to be increased, the *TPC* metric measurement needs to be tweaked. Currently, *TPC* value gives equal weightage to references and citations. An additional weighting factor needs to be multiplied to the citation count. Higher weights can be assigned to help boost specific kinds of papers. This strategy was used for boosting grey literature references in the case study (refer Appendix D). As highlighted earlier in Section 5.3.1.2 of Chapter 5, the AKR technique and *TPC* value is based on author-specified keywords metadata of research papers. Most research paper datasets such as CiteSeer, Microsoft Academic Graph and ACL anthology have not included this field thus far. Alternatively, scientific key term extraction techniques (Lopez & Romary, 2010) can be used to identify topics from the abstract and full text of research papers. Subsequently, these extracted terms could be used for *TPC* calculation.

The IDSP (Integrated Discovery of Similar Papers) technique from Task 2 is a hybrid technique that integrates existing textual and non-textual paper discovery techniques with a new technique based on the features textual similarity and specificity. The current design of IDSP technique gives equal preference to all input papers. There is a possibility that a paper with a large number of citations, could

impact the recommendation process as the resultant papers might be closer to such a paper than the other papers. This could be considered as a minor limitation. A normalizing factor can be added to each paper in the IDSP technique so that all papers exert equal influence in the recommendation process.

The CNS technique of Task 3 provides the most scope for improvements among the three tasks since it is a novel task in the context of SPRS studies. For instance, the recommendation mix can be diversified based on the article-type preference of the user. The lack of diversity in the recommendations of Task 3 could be considered as a minor limitation. As a mitigatory exercise, a precursory in-depth citation analysis can be conducted with the back-end dataset so that the recommendation mix can be identified for each of the chosen article-types.

The recommendation models in the Rec4LRW system are non-personalized in lieu of conventional RS studies. The only mode of input in the Rec4LRW system is explicitly handled by the users in the form of research topic (Task 1) and research papers (Tasks 2 and 3). On the other hand, personalized RS models are largely based on implicit data items such as user's session logs and publication history. However, certain participants in the user evaluation study wanted personalized recommendations instead of general experience-agnostic recommendations. Both the count and recency of papers in the Rec4LRW corpus had a negative impact on the user experience for some participants. This was a limitation explicitly stated in the user guide provided to the participants. Secondly, papers from popular sources such as IEEE Xplore were not made available in the system.

During the user evaluation study, the participants had to select one research topic from the list of 43 available topics in Task 1. This handicapped the researchers in

expressing a research topic of their choice through search keywords. This issue could have been avoided if the corpus had a bigger quantity of papers collated from more sources. The impact of the informational display features has been ascertained solely through the participants' subjective feedback data. The lack of control vs. experimental group comparison limits the current claim that these features positively impacted the user experience. In the study design of Task 2, the number of papers in the seed basket was arbitrarily set to five on the grounds that users could witness the real potential of the recommendation task. In real-world settings, researchers might search for similar papers based on a lesser number of seed papers, as indicated by some participants. Hence, this could be seen as a limitation. During Task 3 evaluation, participants were requested to position themselves in a hypothetical scenario where they have completed LR, data collection, data analysis and they were about to write papers for publication purposes. In such a scenario, Task 3 was supposed to assist the participants in shortlisting papers from the final reading list. Even though, the evaluation procedures are based on instant self-reporting of the perceived measures, Task 3 evaluation could have been better served through a longitudinal study.

6.6 Recommendations for Future Work

There are two phases of future studies to build on the initial base setup by the first evaluation study conducted with the Rec4LRW system. In the first phase, the Rec4LRW system will be tested with the system customization and user personalization aspects. The aspects that are to be considered from system customization are UI-features (sort options, facets and advanced search), control feature on recommendations count and Bibtex file submission feature. User personalization aspects that are to be considered are favorites specification and paper anchors. In Task 2, the participants will be allowed to select a minimum of two papers

instead of the five papers in the current user evaluation study. Free-text searching would be enabled in Task 1 so that participants could enter the search keywords.

In the second phase of Rec4LRW user studies, algorithmic independence and relevance feedback based mechanisms would be tested. Metrics from other similar systems would be incorporated to provide more information about the recommended papers. Cross-linking with systems such as Semantic Scholar, Aminer and CiteSeerX would be enabled and testing would be performed to understand whether the feature is useful and effective for researchers.

In addition to the standard way of evaluating SPRS through a single-session evaluation study, longitudinal user studies are to be conducted for two reasons. The first reason is the intent to validate the performance of the system in meeting the user expectations over an extended period of time. Secondly, such an evaluation approach would help in identifying unknown user behaviors and also new user requirements. The main impediment to such an evaluation strategy is the dataset/corpus used in the system. Most of the SPRS implementations are built on top of datasets which comprise of historic papers and not the most recently published papers. This issue could be potentially mitigated by incorporating live results from popular information sources such as Google Scholar so that user interest is maintained within the system. For the Rec4LRW system, there are plans to allow the university participants to use the system for a longer period of time.

The quality of SPRS recommendations can be further improved if recommendations are formulated based on citations motivation data. Erikson & Erlandson (2014) put forth a taxonomy of motivations based the citation intent of researchers. The main categories are argumentation (support for claims), social

alignment (pushing author's identity), mercantile alignment (uplifting other researchers' works) and data (evidence). Other citation motivation studies (Case & Higgins, 2000; Case & Miller, 2011; Harwood, 2009) have reported along similar lines. A study has already been conducted on a similar theme (Liu, Yu, Guo, Sun, & Gao, 2014). However, the citation motivations are not as detailed as expected; instead they are based on topics. If the citation motivations are obtained on a detailed level, then highly accurate recommendations can be made to users depending on the context of use. The citation motivation data can be expressed as a microdata in the recently developed Research Articles in Simplified HTML (RASH) markup language (Di Iorio et al., 2016) through which scientific articles could be written in a coded manner. RASH is useful both for data extraction in the future and also for automatic manuscript conversion from one style to another (e.g. Springer LNCS to ACM).

It has been highlighted earlier that datasets usage in SPRS is an issue as there are very less standardized datasets where citations are explicitly present as a separate field. If the bibliographic references of papers are identifiable using distinct URIs, data could be extracted from different dataset for research purposes. The OpenCitations project (Peroni, Dutton, Gray, & Shotton, 2015) has been started for this purpose. As a part of this project, PubMed citations are being released in RDF format so that the benefits of Linked Data (Tom Heath & Bizer, 2011) could be reaped. Data from this project should be used for SPRS studies since the reliance on conventional datasets is reduced. Secondly, there will be no need to perform extraction of citations from the full-text of papers. However, the data from this source is currently limited to PubMed papers.

The field of Brain-computer interface (BCI) which was introduced in 1973 (Vidal, 1973), is one of the most interesting fields of research as it deals with methods

for extracting information directly from the human brain. One of the main limitations in IR and RS (to some extent) is the limited ability of expressing human needs through search keywords (Saracevic, 2007a). This limitation is addressed by BCI since the information can be directly retrieved from the brain. The usage of BCI in SPRS has already been initiated by the MindSee project (Gamberini et al., 2015) where there are plans for integrating peripheral physiological data with eye-tracking data so that systems could adapt to users' relevance and satisfaction with search results. If this research reaches a mature state for actual implementation, it could be of utmost benefit to researchers as wading through scientific literature could become a simple activity. Interesting questions arise on how IR and RS algorithms can be tweaked with BCI data.

REFERENCES

- ACM. (2015). The 2012 ACM Computing Classification System. Retrieved August 14, 2015, from <https://www.acm.org/about/class/2012>
- Agarwal, N., Haque, E., Liu, H., & Parsons, L. (2004). A Subspace Clustering Framework for Research Group Collaboration. *International Journal of Information Technology and Web Engineering (IJITWE)*, 1(1), 35–58. <https://doi.org/10.4018/jitwe.2006010102>
- Agarwal, N., Haque, E., Liu, H., & Parsons, L. (2005). Research Paper Recommender Systems: A Subspace Clustering Approach. In *Advances in Web-Age Information Management* (Vol. 301, pp. 475–491). https://doi.org/10.1007/11563952_42
- Agrawal, R., Imieliński, T., Swami, A., Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93* (Vol. 22, pp. 207–216). New York, New York, USA: ACM Press. <https://doi.org/10.1145/170035.170072>
- Altman, D. G. (1990). *Practical Statistics for Medical Research*. Chapman and Hall/CRC.
- AnyStyle. (2015). AnyStyle.io. Retrieved July 22, 2015, from <http://anystyle.io/>
- Atanassova, I., & Bertin, M. (2014). Faceted Semantic Search for Scientific Papers. In *The Semantic Publishing Challenge, Task 3: In-use task, 11th European Semantic Web Conference (ESWC-2014)*. Crete, Greece.
- Athukorala, K., Hoggan, E., Lehtio, A., Ruotsalo, T., & Jacucci, G. (2013). Information-Seeking Behaviors of Computer Scientists: Challenges for Electronic

- Literature Search Tools. *Proceedings of the American Society for Information Science and Technology*, 50(1), 1–11. <https://doi.org/10.1002/meet.14505001041>
- Bae, D.-H., Hwang, S.-M., Kim, S.-W., & Faloutsos, C. (2014). On Constructing Seminal Paper Genealogy. *IEEE Transactions on Cybernetics*, 44(1), 54–65. <https://doi.org/10.1109/TCYB.2013.2246565>
- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. ACM Press.
- Balakrishnan, V., Ahmadi, K., & Ravana, S. D. (2014). Improving retrieval relevance using users' explicit feedback. *Aslib Journal of Information Management*, 68(1), 76–98. <https://doi.org/10.1108/AJIM-07-2015-0106>
- Bansal, T., Belanger, D., & McCallum, A. (2016). Ask the GRU: Multi-task Learning for Deep Text Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16* (pp. 107–114). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2959100.2959180>
- Barrett, A. (2005). The Information-Seeking Habits of Graduate Student Researchers in the Humanities. *The Journal of Academic Librarianship*, 31(4), 324–331. <https://doi.org/10.1016/j.acalib.2005.04.005>
- Basu, C., Hirsh, H., Cohen, W. W., & Nevill-manning, C. (2001). Technical Paper Recommendation : A Study in Combining Multiple Information Sources. *Journal of Artificial Intelligence Research*, 1, 231–252. <https://doi.org/10.1613/jair.739>
- Bates, M. J. (2002). Toward an integrated model of information seeking and searching. *New Review of Information Behaviour Research*, 3, 1–15.
- Bates, M. J. (2005). An Introduction to Metatheories, Theories and Models. In K. . Fisher, S. Erdelez, & L. Mckechnie (Eds.), *Theories of Information Behavior* (1st

- ed., pp. 1–24). Information Today Inc.
- Beel, J. (2015). *Towards Effective Research-Paper Recommender Systems and User Modeling based on Mind Maps*. Retrieved from <https://arxiv.org/pdf/1703.09109>
- Beel, J., Genzmehr, M., Langer, S., Nürnberger, A., & Gipp, B. (2013). A comparative analysis of offline and online evaluations and discussion of research paper recommender system evaluation. In *Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation - RepSys '13* (pp. 7–14). New York, New York, USA: ACM Press.
<https://doi.org/10.1145/2532508.2532511>
- Beel, J., & Langer, S. (2016). Research Paper Recommender Systems: A Literature Survey. *International Journal on Digital Libraries*, 17(4), 305–338.
<https://doi.org/10.1007/s00799-015-0156-0>
- Beel, J., Langer, S., Genzmehr, M., Gipp, B., Breitingner, C., & Nurnberger, A. (2013). Research Paper Recommender System Evaluation : A Quantitative Literature Survey. In *Proceedings of the Workshop on Reproducibility and Replication in Recommender Systems Evaluation (RepSys) at the ACM Recommender System conference* (pp. 15–22). Hong Kong, China.
<https://doi.org/10.1145/2532508.2532512>
- Beel, J., Langer, S., Genzmehr, M., & Nürnberger, A. (2013). Introducing Docear's Research Paper Recommender System. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 459–460). Indianapolis, Indiana, USA. <https://doi.org/10.1145/2467696.2467786>
- Beierle, F., Tan, J., & Grunert, K. (2016). Analyzing Social Relations for Recommending Academic Conferences. *Proceedings of the 8th ACM*

International Workshop on Hot Topics in Planet-Scale mObile Computing and Online Social neTworking - HotPOST '16, 37–42.

<https://doi.org/10.1145/2944789.2944871>

Belkin, N., & Croft, W. (1992). Information filtering and information retrieval: two sides of the same coin? *Communications of the ACM*, 35(12), 29–38.

<https://doi.org/10.1145/138859.138861>

Berger, A., & Lafferty, J. (1999). Information retrieval as statistical translation. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 222–229). Berkeley, California, USA.

<https://doi.org/10.1145/312624.312681>

Bethard, S., & Jurafsky, D. (2010). Who Should I Cite ? Learning Literature Search Models from Citation Behavior Categories and Subject Descriptors. In C. Toronto, ON (Ed.), *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 609–618).

<https://doi.org/10.1145/1871437.1871517>

Bhavnani, S. K. (2002). Domain-specific search strategies for the effective retrieval of healthcare and shopping information. In *CHI '02 extended abstracts on Human factors in computing systems - CHI '02* (p. 610). New York, New York, USA: ACM Press.

<https://doi.org/10.1145/506443.506508>

Bichteler, J., & Ward, D. (1989). Information-Seeking Behavior of Geoscientists. *Special Libraries*, 80(3), 1691–78.

Björk, B.-C., & Hedlund, T. (2003). Scientific Publication Life-Cycle Model (SPLC).

Retrieved February 25, 2014, from

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.106.4255&rep=rep1&t>

ype=pdf

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993--1022. Retrieved from <http://dl.acm.org/citation.cfm?id=944937>
- Blummer, B., Watulak, S. L., & Kenton, J. (2012). The Research Experience for Education Graduate Students: A Phenomenographic Study. *Internet Reference Services Quarterly*, 17(3-4), 117-146. <https://doi.org/10.1080/10875301.2012.747462>
- Bogers, T., & Bosch, A. van den. (2008). Recommending Scientific Articles Using CiteULike. In *Proceedings of the 2008 ACM conference on Recommender systems* (pp. 287-290). Lausanne, Switzerland. <https://doi.org/10.1145/1454008.1454053>
- Bornmann, L., Weymuth, C., & Daniel, H.-D. (2009). A content analysis of referees' comments: how do comments on manuscripts rejected by a high-impact journal and later published in either a low- or high-impact journal differ? *Scientometrics*, 83(2), 493-506. <https://doi.org/10.1007/s11192-009-0011-4>
- Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, 21(3), 487-508. <https://doi.org/10.1016/j.chb.2004.10.005>
- Brew, C., & McKelvie, D. (1996). Word-pair extraction for lexicography. In *Proceedings of the 2nd International Conference on New Methods in Language Processing* (pp. 45-55). Retrieved from https://www.researchgate.net/profile/Chris_Brew/publication/2431964_Word-Pair_Extraction_for_Lexicography/links/00b7d536db4435e2d2000000.pdf

- Bullock, S. M. (2013). Using digital technologies to support Self-Directed Learning for preservice teacher education. *Curriculum Journal*, 24(1), 103–120.
<https://doi.org/10.1080/09585176.2012.744695>
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331–370.
<https://doi.org/10.1023/A:1021240730564>
- Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management*, 31(2), 191–213.
[https://doi.org/10.1016/0306-4573\(95\)80035-R](https://doi.org/10.1016/0306-4573(95)80035-R)
- Case, D. O., & Higgins, G. M. (2000). How Can We Investigate Citation Behavior ? A Study of Reasons for Citing Literature in Communication. *Journal of the American Society for Information Science*, 51(7), 635–645.
[https://doi.org/10.1002/\(SICI\)1097-4571\(2000\)51:7<635::AID-ASI6>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1097-4571(2000)51:7<635::AID-ASI6>3.0.CO;2-H)
- Case, D. O., & Miller, J. B. (2011). Do Bibliometricians Cite Differently From Other Scholars? *Journal of the American Society for Information Science and Technology*, 62(3), 421–432. <https://doi.org/10.1002/asi.21466>
- Castells, P., Vargas, S., & Wang, J. (2011). Novelty and Diversity Metrics for Recommender Systems: Choice, Discovery and Relevance. Retrieved from <http://hdl.handle.net/10486/666094>
- Catalano, A. (2013). Patterns of graduate students' information seeking behavior: A meta-synthesis of the literature. *Journal Of Documentation*, 9(2), 243–274.
<https://doi.org/10.1108/00220411311300066>
- Chakraborty, T., Krishna, A., Singh, M., Ganguly, N., Goyal, P., & Mukherjee, A.

- (2016). FeRoSA : A Faceted Recommendation System for Scientific Articles. In *Pacific Asia Knowledge Discovery and Data Mining Conference (PAKDD) 2016* (pp. 528–541). Auckland, New Zealand. https://doi.org/10.1007/978-3-319-31750-2_42
- Champiri, Z. D., Shahamiri, S. R., & Salim, S. S. B. (2015). A systematic review of scholar context-aware recommender systems. *Expert Systems with Applications*, 42(3), 1743–1758. <https://doi.org/10.1016/j.eswa.2014.09.017>
- Chandrasekaran, K., Gauch, S., Lakkaraju, P., & Luong, H. P. (2008). Concept-Based Document Recommendations for CiteSeer Authors. In *Adaptive Hypermedia and Adaptive Web-Based Systems* (pp. 83–92). Hannover, Germany: Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-70987-9_11
- Chen, C.-H., Mayanglambam, S. D., Hsu, F.-Y., Lu, C.-Y., Lee, H.-M., & Ho, J.-M. (2011). Novelty Paper Recommendation Using Citation Authority Diffusion. In *Technologies and Applications of Artificial Intelligence (TAAI), 2011 International Conference on* (pp. 126–131). IEEE. <https://doi.org/10.1109/TAAI.2011.29>
- Chen, C., Mao, C., Tang, Y., Chen, G., & Zheng, J. (2013). Personalized Recommendation Based on Implicit Social Network of Researchers. In *Pervasive Computing and the Networked World* (pp. 97–107). Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37015-1_9
- Chen, J., Chen, T., Guo, H., Yu, T., & Wang, W. (2012). Context-aware Document Recommendation by Mining Sequential Access Data. In *Proceedings of the 1st International Workshop on Context Discovery and Data Mining - ContextDD '12* (p. 6). New York, New York, USA: ACM Press.

<https://doi.org/10.1145/2346604.2346612>

Chen, Z., Xia, F., Jiang, H., Liu, H., & Zhang, J. (2015). Aver: Random Walk Based Academic Venue Recommendations. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 579–584).

<https://doi.org/10.1145/2740908.2741738>

Chou, J.-K., & Yang, C.-K. (2011). PaperVis: Literature Review Made Easy. *Computer Graphics Forum*, 30(3), 721–730. <https://doi.org/10.1111/j.1467-8659.2011.01921.x>

Chowdhury, S., Gibb, F., & Landoni, M. (2014). A model of uncertainty and its relation to information seeking and retrieval (IS&R). *Journal of Documentation*, 70(4), 575–604. <https://doi.org/10.1108/JD-05-2013-0060>

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (Third). Routledge.

Collins. (2016). Definition of “reading list” | Collins English Dictionary. Retrieved February 18, 2016, from <http://www.collinsdictionary.com/dictionary/english/reading-list>

Costas, R., Zahedi, Z., & Wouters, P. (2015). Do “Altmetrics” Correlate With Citations? Extensive Comparison of Altmetric Indicators With Citations From a Multidisciplinary Perspective. *Journal of the Association for Information Science and Technology*, 66(10), 2003–2019. <https://doi.org/10.1002/asi.23309>

Cothran, T. (2011). Google Scholar acceptance and use among graduate students: A

- quantitative study. *Library & Information Science Research*, 33(4), 293–301.
<https://doi.org/10.1016/j.lisr.2011.02.001>
- Curró, G., & Pretto, G. (2016). An approach for doctoral students conducting a context-specific review of literature in IT, ICT and Educational Technology. *New Review of Academic Librarianship*, 00–00.
<https://doi.org/10.1080/13614533.2016.1227861>
- Debachere, M.-C. (1995). Problems in obtaining grey literature. *IFLA Journal*, 21(2), 94–98. <https://doi.org/10.1177/034003529502100205>
- Dehghani, Z., Afshar, E., Jamali, H. R., & Nematbakhsh, M. A. (2011). A multi-layer contextual model for recommender systems in digital libraries. *Aslib Proceedings*, 63(6), 555–569. <https://doi.org/10.1108/00012531111187216>
- Derntl, M. (2014). Basics of research paper writing and publishing. *International Journal of Technology Enhanced Learning*, 6(2), 105–123.
<https://doi.org/10.1504/ijtel.2014.066856>
- Dey, I. (1993). *Qualitative Data Analysis: A User-friendly Guide for Social Scientists*. New York, NY: Routledge. Retrieved from
https://books.google.com.sg/books/about/Qualitative_Data_Analysis.html?id=_Cl dPVi2g1cC&redir_esc=y
- Di Cesare, R., Luzi, D., & National, I. (2007). The impact of Grey Literature in the web environment: A citation analysis using Google Scholar. In *Proceedings of the Nineth International Conference on Grey Literature* (Vol. 4, pp. 49–63). Antwerp, Belgium. Retrieved from
http://www.textrelease.com/images/GL9_Conference_Proceedings.pdf#page=49
- Di Iorio, A., Gonzalez-Beltran, A., Osborne, F., Peroni, S., Poggi, F., & Vitali, F.

- (2016). It ROCS!: The RASH Online Conversion Service. In *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 25–26). Montréal, Québec, Canada: International World Wide Web Conferences Steering Committee. <https://doi.org/10.1145/2872518.2889408>
- Ding, Y., Chowdhury, G. G., Foo, S., & Qian, W. (2000). Bibliometric information retrieval system (BIRS): A web search interface utilizing bibliometric research results. *Journal of the American Society for Information Science*, 51(13), 1190–1204. [https://doi.org/10.1002/1097-4571\(2000\)9999:9999<::AID-ASII1031>3.0.CO;2-B](https://doi.org/10.1002/1097-4571(2000)9999:9999<::AID-ASII1031>3.0.CO;2-B)
- Diriye, A., Blandford, A., Tombros, A., & Vakkari, P. (2013). The Role of Search Interface Features During Information Seeking. In *Research and Advanced Technology for Digital Libraries* (pp. 235–240). https://doi.org/10.1007/978-3-642-40501-3_23
- Donohew, L., Tipton, L., & Haney, R. (1978). Analysis of Information-Seeking Strategies. *Journalism Quarterly*, 55(1), 25–31. <https://doi.org/10.1177/107769907805500103>
- Du, J. T., & Evans, N. (2011a). Academic Library Services Support For Research Information Seeking. *Australian Academic & Research Libraries*, 42(2), 103–120. <https://doi.org/10.1080/00048623.2011.10722217>
- Du, J. T., & Evans, N. (2011b). Academic Users' Information Searching on Research Topics: Characteristics of Research Tasks and Search Strategies. *The Journal of Academic Librarianship*, 37(4), 299–306. <https://doi.org/10.1016/j.acalib.2011.04.003>
- Dwork, C., Kumar, R., Naor, M., & Sivakumar, D. (2001). Rank aggregation methods

- for the web. In *Proceedings of the 10th international conference on World Wide Web* (pp. 613–622). Hong Kong, Hong Kong.
<https://doi.org/10.1145/371920.372165>
- Eirinaki, M., & Vazirgiannis, M. (2003). Web mining for web personalization. *ACM Transactions on Internet Technology (TOIT)*, 3(1), 1–27.
<https://doi.org/10.1145/643477.643478>
- Ekstrand, M. D., Kannan, P., Stemper, J. A., Butler, J. T., Konstan, J. A., & Riedl, J. T. (2010). Automatically Building Research Reading Lists. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 159–166). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1864708.1864740>
- Ellis, D. (1993). Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *The Library Quarterly*, 63(4), 469–486.
<https://doi.org/10.1086/602622>
- Ellis, D., Cox, D., & Hall, K. (1993). A Comparison of the Information Seeking Patterns of Researchers in the Physical and Social Sciences. *Journal of Documentation*, 49(4), 356–369. <https://doi.org/10.1108/eb026919>
- Ellis, D., & Haugan, M. (1997). Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation*, 53(4), 384–403. <https://doi.org/10.1108/eum0000000007204>
- Eppler, M., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society*, 20(5), 325–344.
<https://doi.org/10.1080/01972240490507974>
- Erikson, M. G., & Erlandson, P. (2014). A taxonomy of motives to cite. *Social Studies*

- of Science*, 44(4), 625–637. <https://doi.org/10.1177/0306312714522871>
- Fidzani, B. T. (1998). Information needs and information-seeking behaviour of graduate students at the University of Botswana. *Library Review*, 47(7), 329–340. <https://doi.org/10.1108/00242539810233459>
- Forrester, J. W. (1971). Counterintuitive behavior of social systems. *Technological Forecasting and Social Change*, 3, 1–22. [https://doi.org/10.1016/S0040-1625\(71\)80001-X](https://doi.org/10.1016/S0040-1625(71)80001-X)
- Gamberini, L., Spagnolli, A., Blankertz, B., Kaski, S., Freeman, J., Acqualagna, L., ... Jacucci, G. (2015). Developing a Symbiotic System for Scientific Information Seeking: The MindSee Project. In B. Blankertz, G. Jacucci, L. Gamberini, A. Spagnolli, & J. Freeman (Eds.), *Symbiotic Interaction: 4th International Workshop, Symbiotic 2015, Berlin, Germany, October 7--8, 2015, Proceedings* (pp. 68–80). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-24917-9_7
- Garfield, E. (2003). The meaning of the impact factor. *International Journal of Clinical and Health Psychology*, 3(2), 363–369. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=498807>
- Gentner, D., & Stevens, A. L. (1983). *Mental Models*. New Jersey, USA: Lawrence Erlbaum Associates, Publishers. Retrieved from https://books.google.com.sg/books/about/Mental_Models.html?id=G8iYAgAAQBAJ&redir_esc=y
- George, C. A., Bright, A., Hurlbert, T., Linke, E. C., Clair, G. S., & Stein, J. (2006). Scholarly Use of Information: Graduate Students' Information Seeking Behaviour. *Information Research*, 11(4). Retrieved from

<http://www.informationr.net/ir/11-4/paper272.html>

Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998). CiteSeer: An Automatic Citation Indexing System. In *Proceedings of the third ACM conference on Digital libraries* (pp. 89–98). Pittsburgh, Pennsylvania, USA.

<https://doi.org/10.1145/276675.276685>

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), 7821–7826. <https://doi.org/10.1073/pnas.122653799>

Giustini, D., & Boulos, M. N. K. (2013). Google Scholar is not enough to be used alone for systematic reviews. *Online Journal of Public Health Informatics*, 5(2).

<https://doi.org/10.5210/ojphi.v5i2.4623>

Gokhale, D. V., & Kullback, S. (1978). *The Information in Contingency Tables*.

Marcel Dekker Incorporated. Retrieved from

https://books.google.com.sg/books/about/The_Information_in_Contingency_Tables.html?id=oifMmgEACAAJ&redir_esc=y

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61–70. <https://doi.org/10.1145/138859.138867>

Gonçalves, M. A., Fox, E. A., Watson, L. T., & Kipp, N. A. (2004). Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries. *ACM Transactions on Information Systems*, 22(2), 270–312.

<https://doi.org/10.1145/984321.984325>

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2), 337–356. Retrieved from

<http://aisel.aisnet.org/misq/vol37/iss2/3>

Guan, Z., Wang, C., Bu, J., Chen, C., Yang, K., Cai, D., & He, X. (2010). Document recommendation in social tagging services. In *Proceedings of the 19th international conference on World wide web* (pp. 391–400). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1772690.1772731>

Gunawardena, S. D. (2013). *Recommending Research Profiles for Multidisciplinary Academic Collaboration*. Retrieved from <http://dspace.library.drexel.edu/handle/1860/4180>

Gunawardena, S. D., & Weber, R. (2009). Discovering Patterns of Collaboration for Recommendation. In *Proceedings 22th International FLAIRS Conference, FLAIRS'09. AAAI Press, Menlo Park, California*. Retrieved from <http://www.aaai.org/ocs/index.php/FLAIRS/2009/paper/download/121</315%5Cnhttp://hdl.handle.net/1860/4145>

Gupta, M., Li, R., Yin, Z., & Han, J. (2010). Survey on social tagging techniques. *ACM SIGKDD Explorations Newsletter*, 12(1), 58. <https://doi.org/10.1145/1882471.1882480>

Hahn, S.-H. (2016). Technology foresight through the collaboration with human expert and machine intelligence. In *2016 8th International Conference on Knowledge and Smart Technology (KST)* (pp. XVIII–XVIII). IEEE. <https://doi.org/10.1109/KST.2016.7440470>

Han, L., Kashyap, A., Finin, T., Mayfield, J., & Weese, J. (2012). UMBC EBIQUITY-CORE: Semantic Textual Similarity Systems. In *Proceedings of the Second Joint Conference on Lexical and Computational Semantics* (pp. 16–33). Retrieved from <http://ebiquity.umbc.edu/papers/select/person/tim/finin/>

- Hanani, U., Shapira, B., & Shoval, P. (2001). Information Filtering: Overview of Issues, Research and Systems. *User Modeling and User-Adapted Interaction*, 11(3), 203–259. <https://doi.org/10.1023/A:1011196000674>
- Hannon, J., Bennett, M., & Smyth, B. (2010). Recommending twitter users to follow using content and collaborative filtering approaches. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 199–206). <https://doi.org/10.1145/1864708.1864746>
- Harley, D., Acord, S. K., Earl-Novell, S., & Lawrence, S. (2010). Assessing the future landscape of scholarly communication: An exploration of faculty values and needs in seven disciplines. *Center for Studies in Higher Education*. Retrieved from <https://escholarship.org/uc/item/15x7385g.pdf>
- Harwood, N. (2009). An interview-based study of the functions of citations in academic writing across two disciplines. *Journal of Pragmatics*, 41(3), 497–518. <https://doi.org/10.1016/j.pragma.2008.06.001>
- Hausenblas, M. (2009). Exploiting linked data to build web applications. *IEEE Internet Computing*, 13(4), 68–73. <https://doi.org/10.1109/mic.2009.79>
- He, Q., Kifer, D., Pei, J., Mitra, P., & Giles, C. L. (2011). Citation recommendation without author supervision. In *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11* (pp. 755–764). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1935826.1935926>
- He, Q., Pei, J., Kifer, D., Mitra, P., & Giles, L. (2010). Context-aware citation recommendation. *Proceedings of the 19th International Conference on World Wide Web - WWW '10*, 421–430. <https://doi.org/10.1145/1772690.1772734>
- Head, A. J. (2007). Beyond Google: How do students conduct academic research?

- First Monday*, 12(8). <https://doi.org/10.5210/fm.v12i8.1998>
- Heath, T. (2002). A Quantitative Analysis of PhD Students' Views of Supervision. *Higher Education Research & Development*, 21(1), 41–53.
<https://doi.org/10.1080/07294360220124648>
- Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1), 1–136. <https://doi.org/10.2200/S00334ED1V01Y201102WBE001>
- Herrera, G. (2011). Google Scholar Users and User Behaviors: An Exploratory Study. *College & Research Libraries*, 72(4), 316–330. <https://doi.org/10.5860/crl-125rl>
- Hjørland, B. (2013). Citation analysis: A social and dynamic approach to knowledge organization. *Information Processing & Management*, 49(6), 1313–1325.
<https://doi.org/http://dx.doi.org/10.1016/j.ipm.2013.07.001>
- Hoeber, O., & Khazaei, T. (2015). Evaluating citation visualization and exploration methods for supporting academic search tasks. *Online Information Review*, 39(2), 229–254. <https://doi.org/10.1108/OIR-10-2014-0259>
- Holmes, D., & McCabe, C. M. (2002). Improving precision and recall for soundex retrieval. In *Proceedings of International Conference on Information Technology: Coding and Computing* (pp. 22–26). <https://doi.org/10.1109/ITCC.2002.1000354>
- Huang, W., Kataria, S., Caragea, C., Mitra, P., Giles, C. L., & Rokach, L. (2012). Recommending Citations : Translating Papers into References. In *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 1910–1914). Maui, Hawaii, USA.
<https://doi.org/10.1145/2396761.2398542>
- Huang, W., Wu, Z., Mitra, P., & Giles, C. L. (2014). RefSeer : A Citation

- Recommendation System. In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 371–374). London, UK.
<https://doi.org/10.1109/jcdl.2014.6970192>
- Hurtado Martín, G., Schockaert, S., Cornelis, C., & Naessens, H. (2013). Using semi-structured data for assessing research paper similarity. *Information Sciences*, 221, 245–261. <https://doi.org/10.1016/j.ins.2012.09.044>
- Huurdeman, H. C., & Kamps, J. (2014). From multistage information-seeking models to multistage search systems. In *Proceedings of the 5th Information Interaction in Context Symposium on - IliX '14* (pp. 145–154). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2637002.2637020>
- Huurdeman, H. C., Wilson, M. L., & Kamps, J. (2016). Active and Passive Utility of Search Interface Features in Different Information Seeking Task Stages. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval - CHIIR '16* (pp. 3–12). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2854946.2854957>
- Huynh, T., & Hoang, K. (2012). Modeling Collaborative Knowledge of Publishing Activities for Research Recommendation. In *Computational Collective Intelligence. Technologies and Applications* (pp. 41–50).
https://doi.org/10.1007/978-3-642-34630-9_5
- Huynh, T., Hoang, K., Do, L., Tran, H., Luong, H., & Gauch, S. (2012). Scientific publication recommendations based on collaborative citation networks. In *2012 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 316–321). IEEE. <https://doi.org/10.1109/CTS.2012.6261069>
- Hwang, S.-Y., Wei, C.-P., & Liao, Y.-F. (2010). Coauthorship networks and academic

- literature recommendation. *Electronic Commerce Research and Applications*, 9(4), 323–334. <https://doi.org/10.1016/j.elerap.2010.01.001>
- Ingwersen, P., & Järvelin, K. (2006). *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer Science & Business Media. Retrieved from <https://books.google.com/books?hl=en&lr=&id=zuptrfHytHMC&pgis=1>
- Ishimura, Y., & Bartlett, J. C. (2014). Are Librarians Equipped to Teach International Students? A Survey of Current Practices and Recommendations for Training. *Journal of Academic Librarianship*, 40(3), 313–321. <https://doi.org/10.1016/j.acalib.2014.04.009>
- Jaccard, P. (1912). THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1. *New Phytologist*, 11(2), 37–50. <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>
- Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). *Recommender systems: an introduction*. Cambridge University Press.
- Jardine, J. G. (2014). *Automatically generating reading lists*. Retrieved from <https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-848.pdf>
- Jeh, G., & Widom, J. (2002). SimRank : A Measure of Structural-Context Similarity. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 538–543). Edmonton, Alberta, Canad. <https://doi.org/10.1145/775107.775126>
- Jiang, Y., Jia, A., Feng, Y., & Zhao, D. (2012). Recommending academic papers via users' reading purposes. *Proceedings of the Sixth ACM Conference on Recommender Systems - RecSys '12*, 241. <https://doi.org/10.1145/2365952.2366004>

- John E. Hocking, Don W. Stacks, & Steven T. McDermott. (2003). *Communication Research* (Third). Boston: Pearson Education.
- Johnson, J. D. (1997). *Cancer-related information seeking*. Hampton Press.
- Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1), 11–21.
<https://doi.org/10.1108/eb026526>
- Jones, K. S., Walker, S., & Robertson, S. E. (2000). A probabilistic model of information retrieval: development and comparative experiments: Part 2. *Information Processing & Management*, 36(6), 809–840.
[https://doi.org/10.1016/S0306-4573\(00\)00016-9](https://doi.org/10.1016/S0306-4573(00)00016-9)
- Kahlor, L. (2010). PRISM: A Planned Risk Information Seeking Model. *Health Communication*, 25(4), 345–356. <https://doi.org/10.1080/10410231003775172>
- Karlsson, L., Koivula, L., Ruokonen, I., Kajaani, P., Antikainen, L., & Ruismäki, H. (2012). From Novice to Expert: Information Seeking Processes of University Students and Researchers. *Procedia - Social and Behavioral Sciences*, 45, 577–587. <https://doi.org/10.1016/j.sbspro.2012.06.595>
- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25. <https://doi.org/10.1002/asi.5090140103>
- Kim, Y., Seo, J., Croft, W. B., & Smith, D. A. (2014). Automatic suggestion of phrasal-concept queries for literature search. *Information Processing and Management*, 50(4), 568–583. <https://doi.org/10.1016/j.ipm.2014.03.003>
- Kiryakov, A., Popov, B., Terziev, I., Manov, D., & Ognyanoff, D. (2004). Semantic annotation, indexing, and retrieval. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2(1), 49–79.

<https://doi.org/10.1016/j.websem.2004.07.005>

- Kleinberg, J. M. (1999). Hubs , Authorities , and Communities. *ACM Computing Surveys (CSUR)*, 31(4). <https://doi.org/10.1145/345966.345982>
- Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User Modelling and User-Adapted Interaction*, 22(4–5), 441–504. <https://doi.org/10.1007/s11257-011-9118-4>
- Konstan, J. a., & Riedl, J. (2012). Recommender systems: From algorithms to user experience. *User Modelling and User-Adapted Interaction*, 22(1–2), 101–123. <https://doi.org/10.1007/s11257-011-9112-x>
- Krikelas, J. (1983). Information-Seeking Behavior: Patterns and Concepts. *Drexel Library Quarterly*, 19(2), 5–20. Retrieved from <https://eric.ed.gov/?id=EJ298483>
- Küçükünç, O., Saule, E., Kaya, K., & Çatalyürek, Ü. V. (2015). Diversifying citation recommendations. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(4), 55:1-55:21. <https://doi.org/10.1145/2668106>
- Küçükünç, O., Saule, E., Kaya, K., & Çatalyürek, Ü. V. (2013). TheAdvisor : A Webservice for Academic Recommendation. In *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries* (pp. 433–434). New York, NY, USA. <https://doi.org/10.1145/2467696.2467752>
- Kukich, K. (1992). Technique for automatically correcting words in text. *ACM Computing Surveys*, 24(4), 377–439. <https://doi.org/10.1145/146370.146380>
- Lai, C.-H., & Liu, D.-R. (2009). Integrating knowledge flow mining and collaborative filtering to support document recommendation. *Journal of Systems and Software*, 82(12), 2023–2037. <https://doi.org/10.1016/j.jss.2009.06.044>

- Lakshminarayanan, B. (2010). *Towards developing an integrated model of information behavior*. Queensland University of Technology. Retrieved from http://eprints.qut.edu.au/33252/1/Bhuvaneshwari_Lakshminarayanan_Thesis.pdf
- Lawrence, S., Lee Giles, C., & Bollacker, K. (1999). Digital libraries and autonomous citation indexing. *Computer*, 32(6), 67–71. <https://doi.org/10.1109/2.769447>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>
- Lee, J., Lee, K., & Kim, J. G. (2013). *Personalized Academic Research Paper Recommendation System*. Retrieved from <https://arxiv.org/abs/1304.5457>
- Levy, Y., & Ellis, T. J. (2006). A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. *Informing Science: The International Journal of an Emerging Transdiscipline*, 9(1), 181–212. Retrieved from <http://www.inform.nu/Articles/Vol9/V9p181-212Levy99.pdf>
- Li, Y., & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management*, 44(6), 1822–1837. <https://doi.org/10.1016/j.ipm.2008.07.005>
- Liang, Y., Li, Q., & Qian, T. (2011). Finding Relevant Papers Based on Citation Relations. In *Web-Age Information Management: 12th International Conference, WAIM 2011, Wuhan, China, September 14-16, 2011. Proceedings* (Vol. 6897, pp. 403–414). https://doi.org/10.1007/978-3-642-23535-1_35
- Liu, D., Chin-Hui, L., & Chen, Y.-T. (2012). Document Recommendations Based on Knowledge Flows: A Hybrid of Personalized and Group-Based Approaches. *Journal of the American Society for Information Science and Technology*, 63(10), 2100–2117. <https://doi.org/10.1002/asi>

- Liu, J. (2015). User assessment of search task difficulty: Relationships between reasons and ratings. *Library & Information Science Research*, 37(4), 329–337. <https://doi.org/10.1016/j.lisr.2015.05.003>
- Liu, J., Cole, M. J., Liu, C., Bierig, R., Gwizdka, J., Belkin, N. J., ... Zhang, X. (2010). Search behaviors in different task types. In *Proceedings of the 10th annual joint conference on Digital libraries - JCDL '10* (p. 69). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1816123.1816134>
- Liu, J., & Kim, C. S. (2013). Why Do Users Perceive Search Tasks As Difficult ? Exploring Difficulty in Different Task Types. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*. Vancouver BC, Canada. <https://doi.org/10.1145/2528394.2528399>
- Liu, X., Yu, Y., Guo, C., Sun, Y., & Gao, L. (2014). Full-Text based Context-Rich Heterogeneous Network Mining Approach for Citation Recommendation. In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries* (p. 494). <https://doi.org/10.1109/jcdl.2014.6970191>
- Lopez, P., & Romary, L. (2010). HUMB: Automatic Key Term Extraction from Scientific Articles in GROBID. In *Proceedings of the 5th International Workshop on Semantic Evaluation* (pp. 248–251). Los Angeles, California. Retrieved from <http://dl.acm.org/citation.cfm?id=1859719>
- Lops, P., Gemmis, M. De, & Semeraro, G. (2011). Content-based Recommender Systems: State of the Art and Trends. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender Systems Handbook* (pp. 73–105). Boston, MA: Springer US. <https://doi.org/10.1007/978-0-387-85820-3>
- Lü, L., Medo, M., Yeung, C., & Zhang, Y. (2012). Recommender systems. *Physics*

- Reports*, 519(1), 1–49. <https://doi.org/10.1016/j.physrep.2012.02.006>
- Lu, Y., He, J., Shan, D., & Yan, H. (2011). Recommending citations with translation model. In *Proceedings of the 20th ACM international conference on Information and knowledge management* (pp. 2017–2020). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2063576.2063879>
- MacRoberts, M. H., & MacRoberts, B. R. (1988). Author Motivation for Not Citing Influences: A Methodological Note. *Journal of the American Society for Information Science*, 39(6), 432–433. [https://doi.org/10.1002/\(SICI\)1097-4571\(198811\)39:6<432::AID-ASI8>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1097-4571(198811)39:6<432::AID-ASI8>3.0.CO;2-2)
- MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of Citation Analysis: A Critical Review. *Journal of the American Society for Information Science*, 40(5), 342–349. [https://doi.org/10.1002/\(SICI\)1097-4571\(198909\)40:5<342::AID-ASI7>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-4571(198909)40:5<342::AID-ASI7>3.0.CO;2-U)
- MacRoberts, M. H., & MacRoberts, B. R. (2010). Problems of Citation Analysis: A Study of Uncited and Seldom-Cited Influences. *Journal of the American Society for Information Science and Technology*, 61(1), 1–12. <https://doi.org/10.1002/asi.21228>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge: Cambridge university press. Retrieved from http://www.langtoninfo.co.uk/web_content/9780521865715_frontmatter.pdf
- Manouselis, N., & Costopoulou, C. (2007). Analysis and classification of multi-criteria recommender systems. *World Wide Web*, 10(4), 415–441. Retrieved from <http://link.springer.com/article/10.1007/s11280-007-0019-8>
- Manouselis, N., & Verbert, K. (2013). Layered Evaluation of Multi-Criteria

- Collaborative Filtering for Scientific Paper Recommendation. *Procedia Computer Science*, 18, 1189–1197. <https://doi.org/10.1016/j.procs.2013.05.285>
- Markauskaite, L. (2007). Exploring the structure of trainee teachers' ICT literacy: the main components of, and relationships between, general cognitive and technical capabilities. *Educational Technology Research and Development*, 55(6), 547–572. <https://doi.org/10.1007/s11423-007-9043-8>
- McKercher, B., Law, R., Weber, K., & Hsu, C. (2007). Why Referees Reject Manuscripts. *Journal of Hospitality & Tourism Research*, 31(4), 455–470. <https://doi.org/10.1177/1096348007302355>
- Mcnee, S. M. (2006). *Meeting User Information Needs in Recommender Systems*. Proquest. Retrieved from <http://search.proquest.com/docview/305306133>
- McNee, S. M., Kapoor, N., & Konstan, J. A. (2006). Don't Look Stupid : Avoiding Pitfalls when Recommending Research Papers. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (pp. 171–180). Banff, Alberta, Canada. <https://doi.org/10.1145/1180875.1180903>
- McNee, S. M., Riedl, J., & Konstan, J. A. (2006). Being accurate is not enough. In *CHI '06 extended abstracts on Human factors in computing systems - CHI EA '06* (p. 1097). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1125451.1125659>
- Mutschke, P., & Mayr, P. (2015). Science models for search: a study on combining scholarly information retrieval and scientometrics. *Scientometrics*, 102(3), 2323–2345. <https://doi.org/10.1007/s11192-014-1485-2>
- Myers, J. L., Well, A., & Lorch, R. F. (2010). *Research design and statistical analysis* (Third). Routledge.

- Naak, A., Hage, H., & Aïmeur, E. (2008). Papyres : a Research Paper Management System. In *E-Commerce Technology and the Fifth IEEE Conference on Enterprise Computing, E-Commerce and E-Services, 2008 10th IEEE Conference on* (pp. 201–208). <https://doi.org/10.1109/CEC/EEE.2008.53>
- Naak, A., Hage, H., & Aïmeur, E. (2009). A Multi-criteria Collaborative Filtering Approach for Research Paper Recommendation in Papyres. In *E-Technologies: Innovation in an Open World. MCETECH 2009* (pp. 25–39). Ottawa, Canada. https://doi.org/10.1007/978-3-642-01187-0_3
- Needham, G. (2007). Information Anxiety and African-American Students in a Graduate Education Program. *Evidence Based Library and Information Practice (EBLIP)*, 2(1), 147–148. <https://doi.org/10.18438/b8ww2j>
- Nicholas, D., & Rowlands, I. (2011). Social media use in the research workflow. *Information Services & Use*, 31(1–2), 61–83. <https://doi.org/10.3233/ISU-2011-0623>
- Niu, X., & Hemminger, B. M. (2012). A Study of Factors That Affect the Information-Seeking Behavior of Academic Scientists. *Journal of the American Society for Information Science and Technology*, 63(2), 336–353. <https://doi.org/10.1002/asi.21669>
- O'reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. Retrieved January 27, 2014, from <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>
- Page, L., Brin, S., Motwami, R., & Winograd, T. (1999). *The PageRank Citation Ranking: Bringing Order to the Web*. Retrieved from <http://ilpubs.stanford.edu:8090/422/>

- Palmer, J. (1991). Scientists and information: II. Personal factors in information behaviour. *Journal of Documentation*, 47(3), 254–275.
<https://doi.org/10.1108/eb026880>
- Pan, L., Dai, X., Huang, S., & Chen, J. (2015). Academic Paper Recommendation Based on Heterogeneous Graph. *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, 9427, 381–392.
<https://doi.org/10.1007/978-3-642-41491-6>
- Parra, D. (2009). *RecULike : Recommending Scientific Articles on CiteULike using variations of Collaborative Filtering Algorithms*. Retrieved from <http://www.citeulike.org/user/felixlover4ever/article/8147837>
- Parra, D., & Sahebi, S. (2013). Recommender Systems : Sources of Knowledge and Evaluation Metrics. In *Advanced Techniques in Web Intelligence-2* (pp. 149–175). https://doi.org/10.1007/978-3-642-33326-2_7
- Pasi, G. (2011). Contextual Search : Issues and Challenges. In *Proceedings of the 7th conference on Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society: information Quality in e-Health* (pp. 23–30). Graz, Austria. https://doi.org/10.1007/978-3-642-25364-5_3
- Pasi, G., Informatica, D., & Bordogna, G. (2007). A Multi-Criteria Content-based Filtering System. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 775–776). Amsterdam, The Netherlands. <https://doi.org/10.1145/1277741.1277903>
- Peroni, S., Dutton, A., Gray, T., & Shotton, D. (2015). Setting our bibliographic references free: towards open citation data. *Journal of Documentation*, 71(2), 253–277. <https://doi.org/10.1108/JD-12-2013-0166>

- Pihur, V., & Datta, S. (2009). RankAggreg, an R package for weighted rank aggregation. *BMC Bioinformatics*. Retrieved from <http://www.biomedcentral.com/1471-2105/10/62>
- Resnick, P., & Varian, H. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56–68. <https://doi.org/10.1145/245108.245121>
- Ridley, D. (2012). *The literature review: A step-by-step guide for students*. Sage.
- Robertson, S. E., & Walker, S. (1994). Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 232–241). Springer-Verlag. Retrieved from <http://dl.acm.org/citation.cfm?id=188561>
- Saldana, J. (2009). *The Coding Manual for Qualitative Researchers* (First Edition). London: SAGE.
- Salehi, S., Du, J. T., & Ashman, H. (2015). Examining Personalization in Academic Web Search. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media - HT '15* (pp. 103–111). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2700171.2791039>
- Salkind, N. J. (2010). *Encyclopedia of Research Design*. Sage.
- Salton, G., & McGill, M. J. (1983). *Introduction to Modern Information Retrieval*. McGraw-Hill.
- Saracevic, T. (2007a). Relevance : A Review of the Literature and a Framework for Thinking on the Notion in Information Science . Part II : Nature and Manifestations of Relevance. *Journal of the American Society for Information Science and Technology*, 58(13), 1915–1933. <https://doi.org/10.1002/asi.20682>

- Saracevic, T. (2007b). Relevance : A Review of the Literature and a Framework for Thinking on the Notion in Information Science . Part III : Behavior and Effects of Relevance. *Journal of the American Society for Information Science and Technology*, 58(13), 2126–2144. <https://doi.org/10.1002/asi.20681>
- Savolainen, R. (1995). Everyday life information seeking: Approaching information seeking in the context of “way of life.” *Library & Information Science Research*, 17(3). Retrieved from <http://www.sciencedirect.com/science/article/pii/0740818895900489>
- Savolainen, R. (2015). The interplay of affective and cognitive factors in information seeking and use. *Journal of Documentation*, 71(1), 175–197. <https://doi.org/10.1108/JD-10-2013-0134>
- Schöpfel, J., Stock, C., Farace, D. J., & Frantzen, J. (2005). *Citation analysis and grey literature: Stakeholders in the grey circuit. GL6: Sixth International Conference on Grey Literature*. Retrieved from https://archivesic.ccsd.cnrs.fr/sic_00001534/
- Sesagiri Raamkumar, A., Foo, S., & Pang, N. (2015). Rec4LRW – Scientific Paper Recommender System for Literature Review and Writing. In *Proceedings of the 6th International Conference on Applications of Digital Information and Web Technologies* (pp. 106–119). Hong Kong, China: IOS Press. <https://doi.org/10.3233/978-1-61499-503-6-106>
- Sethi, V., & King, W. R. (1998). *Organizational Transformation Through Business Process Reengineering: Applying the Lessons Learned*. Prentice Hall. Retrieved from https://books.google.com.sg/books/about/Organizational_Transformation_Through_Bu.html?id=a_bkAAAAMAAJ&pgis=1

- She, H.-C., Cheng, M.-T., Li, T.-W., Wang, C.-Y., Chiu, H.-T., Lee, P.-Z., ... Chuang, M.-H. (2012). Web-based undergraduate chemistry problem-solving: The interplay of task performance, domain knowledge and web-searching strategies. *Computers & Education*, 59(2), 750–761.
<https://doi.org/10.1016/j.compedu.2012.02.005>
- Shen, X., Tan, B., & Zhai, C. (2005). Implicit user modeling for personalized search. In *Proceedings of the 14th ACM international conference on Information and knowledge management - CIKM '05* (pp. 824–831). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1099554.1099747>
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the Association for Information Science and Technology*, 24(4), 265–269. <https://doi.org/10.1002/asi.4630240406>
- Snedecor, G. W., & Cochran, W. G. (1989). *Statistical Methods* (eight). Iowa State University Press.
- Song, F., & Croft, W. B. (1999). A general language model for information retrieval. In *Proceedings of the eighth international conference on Information and knowledge management* (pp. 316–321). Kansas City, Missouri, USA.
<https://doi.org/10.1145/319950.320022>
- Speier, C., Valacich, J. S., & Vessey, I. (1999). The Influence of Task Interruption on Individual Decision Making: An Information Overload Perspective. *Decision Sciences*, 30(2), 337–360. <https://doi.org/10.1111/j.1540-5915.1999.tb01613.x>
- Speriosu, M., & Tashiro, T. (2006). *Comparison of Okapi BM25 and Language Modeling Algorithms for NTCIR-6*. Retrieved from
<https://webpace.utexas.edu/mas5622/www/speriosu06.pdf>

- Spezi, V. (2016). Is Information-Seeking Behavior of Doctoral Students Changing?: A Review of the Literature (2010–2015). *New Review of Academic Librarianship*, 1–29. <https://doi.org/10.1080/13614533.2015.1127831>
- Stangor, C. (2014). *Research Methods for the Behavioral Sciences*. Boston, MA: Houghton Mifflin. Retrieved from <https://www.amazon.com/Research-Methods-Behavioral-Sciences-Charles/dp/1285077024>
- Strohman, T., Croft, W. B., & Jensen, D. (2007). Recommending citations for academic papers. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '07* (p. 705). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1277741.1277868>
- Sturm, B., Schneider, S., & Sunyaev, A. (2015). *Leave No Stone Unturned : Introducing a Revolutionary Meta-Search Tool for Rigorous and Efficient Systematic Literature Searches. ECIS 2015 Research-in-Progress Papers*. Münster, Germany. Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1033&context=ecis2015_rip
- Sugiyama, K., & Kan, M.-Y. (2011). Serendipitous Recommendation for Scholarly Papers. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries* (pp. 307–310). Ottawa, Ontario, Canada. <https://doi.org/10.1145/1998076.1998133>
- Sugiyama, K., & Kan, M.-Y. (2013). Exploiting Potential Citation Papers in Scholarly Paper Recommendation. In *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries - JCDL '13* (p. 153). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2467696.2467701>

- Suntharasaj, P., & Kocaoglu, D. F. (2013). Bridging the missing link between “top-down” and “bottom-up”: a case study of Thailand’s international collaboration in S&T. In *2013 Proceedings of PICMET '13: Technology Management in the IT-Driven Services (PICMET)* (pp. 201–219). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/6641627/>
- Switzer, A., & Perdue, S. W. (2011). Dissertation 101: A Research and Writing Intervention for Education Graduate Students. *Education Libraries*, 34(1), 4–14. Retrieved from <http://education.sla.org/wp-content/uploads/2012/12/34-1-4.pdf>
- Tabatabai, D., & Shore, B. M. (2005). How experts and novices search the Web. *Library & Information Science Research*, 27(2), 222–248. <https://doi.org/10.1016/j.lisr.2005.01.005>
- Takeda, H., Veerkamp, P., & Yoshikawa, H. (1990). Modeling Design Process. *AI Magazine*, 11(4). <https://doi.org/10.1609/aimag.v11i4.855>
- Tang, M.-C. (2009). A study of academic library users’ decision-making process: a Lens model approach. *Journal of Documentation*, 65(6), 938–957. <https://doi.org/10.1108/00220410910998933>
- Tang, T. Y., & McCalla, G. (2009). The Pedagogical Value of Papers : a Collaborative-Filtering based Paper Recommender. *Journal of Digital Information*, 10(2). Retrieved from <https://journals.tdl.org/jodi/index.php/jodi/article/view/446>
- Tarjan, R. (1972). Depth-First Search and Linear Graph Algorithms. *SIAM Journal on Computing*, 1(2), 146–160. <https://doi.org/10.1137/0201010>
- Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, 27(2), 237–246.

<https://doi.org/10.1177/1098214005283748>

- Tintarev, N., & Masthoff, J. (2007). A Survey of Explanations in Recommender Systems. In *2007 IEEE 23rd International Conference on Data Engineering Workshop* (pp. 801–810). IEEE. <https://doi.org/10.1109/ICDEW.2007.4401070>
- Torres, R., Mcnee, S. M., Abel, M., Konstan, J. A., & Riedl, J. (2004). Enhancing Digital Libraries with TechLens. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries* (pp. 228–236). <https://doi.org/10.1109/JCDL.2004.240019>
- Vakkari, P. (2001). A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of Documentation*, 57(1), 44–60. <https://doi.org/10.1108/EUM0000000007075>
- Van Deursen, A. J. A. M., & Van Dijk, J. A. G. M. (2009). Using the Internet: Skill related problems in users' online behavior. *Interacting with Computers*, 21(5), 393–402. <https://doi.org/10.1016/j.intcom.2009.06.005>
- van Setten, M., Veenstra, M., Nijholt, A., & van Dijk, B. (2006). Goal-based structuring in recommender systems. *Interacting with Computers*, 18(3), 432–456. <https://doi.org/10.1016/j.intcom.2005.11.005>
- Vargas, S., Hristakeva, M., & Jack, K. (2016). Mendeley: Recommendations for Researchers. In *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16* (pp. 365–365). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2959100.2959116>
- Vellino, A. (2013). *Usage-based vs . Citation-based Methods for Recommending Scholarly Research Articles*. Retrieved from <https://arxiv.org/abs/1303.7149>
- Vellino, A., & Zeber, D. (2007). A Hybrid , Multi-Dimensional Recommender for

- Journal Articles in a Scientific Digital Libraries. In *Proceedings of the 2007 IEEE/WIC/ACM International Conference on Web Intelligence and International Conference on Intelligent Agent Technology* (pp. 111–114).
<https://doi.org/10.1109/WI-IATW.2007.29>
- Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, 39(2), 273–315.
<https://doi.org/10.1111/j.1540-5915.2008.00192.x>
- Verbeke, W., & Ward, R. W. (2006). Consumer interest in information cues denoting quality, traceability and origin: An application of ordered probit models to beef labels. *Food Quality and Preference*, 17(6), 453–467.
<https://doi.org/10.1016/j.foodqual.2005.05.010>
- Vidal, J. J. (1973). Toward Direct Brain-Computer Communication. *Annual Review of Biophysics and Bioengineering*, 2(1), 157–180.
<https://doi.org/10.1146/annurev.bb.02.060173.001105>
- Voorhees, E. M. (1999). The TREC-8 Question Answering Track Report. *TREC*, 99, 77–82. Retrieved from http://trec.nist.gov/pubs/trec8/papers/qa_report.pdf
- Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., ... Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, 5(1), 14–26. <https://doi.org/10.1016/j.joi.2010.06.004>
- Wang, Y., Zhai, E., Hu, J., & Chen, Z. (2010). Claper: Recommend classical papers to beginners. In *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010)* (pp. 2777–2781). IEEE.
<https://doi.org/10.1109/FSKD.2010.5569227>

- Weng, S.-S., & Chang, H.-L. (2008). Using ontology network analysis for research document recommendation. *Expert Systems with Applications*, 34(3), 1857–1869. <https://doi.org/10.1016/j.eswa.2007.02.023>
- Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10), 1550–1560. <https://doi.org/10.1109/5.58337>
- West, J. D. (2010). *Eigenfactor: ranking and mapping scientific knowledge*. PhD Thesis. <https://doi.org/papers2://publication/uuid/949B5484-DBE6-4BCF-BE91-59B6C4F37038>
- West, J. D., Bergstrom, T. C., & Bergstrom, C. T. (2010). The Eigenfactor Metrics™: A Network Approach to Assessing Scholarly Journals. *College & Research Libraries*, 71(3), 236–244. <https://doi.org/10.5860/0710236>
- White, H. D. (1990). Author co-citation analysis: overview and defense. In C. Borgman (Ed.), *Scholarly Communication* (pp. 84–106). Newbury Park, CA: Sage Publications.
- White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32(3), 163–171. <https://doi.org/10.1002/asi.4630320302>
- White, R. W., & Roth, R. A. (2009). Exploratory Search: Beyond the Query-Response Paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1), 1–98. <https://doi.org/10.2200/S00174ED1V01Y200901ICR003>
- Wildemuth, B. M., & Freund, L. (2012). Assigning search tasks designed to elicit exploratory search behaviors. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval - HCIR '12* (pp. 1–10). New York, New York, USA: ACM Press. <https://doi.org/10.1145/2391224.2391228>

- Wilson, M. L. (2011). Search User Interface Design. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 3(3), 1–143.
<https://doi.org/10.2200/S00371ED1V01Y201111ICR020>
- Wilson, P. (1983). *Second-Hand Knowledge: An Inquiry Into Cognitive Authority*. Greenwood Press.
- Wilson, T. (1999). Models in information behaviour research. *Journal of Documentation*, 55(3), 249–270.
- Winoto, P., Tang, T. Y., & Mccalla, G. (2012). Contexts in a Paper Recommendation System with Collaborative Filtering. *The International Review of Research in Open and Distance Learning*, 13(5), 56–75.
<https://doi.org/10.19173/irrodl.v13i5.1243>
- Wolf, S. E., Brush, T., & Saye, J. (2003). Using an Information Problem-Solving Model as a Metacognitive Scaffold for Multimedia-Supported Information-Based Problems. *Journal of Research on Technology in Education*, 35(3), 321–341.
<https://doi.org/10.1080/15391523.2003.10782389>
- Wolfram, D. (2003). *Applied informetrics for information retrieval research*. Greenwood Publishing Group. Retrieved from
[https://books.google.com/books?hl=en&lr=&id=FhPeRRKGLAUC&oi=fnd&pg=PR9&dq=Wolfram,+D.:+Applied+Informetrics+for+Information+Retrieval+Research.+Libraries+Unlimited+\(2003\)&ots=rUtg4fV_r9&sig=UgDKr1D9OQETbnyCt9fF6yRRELc](https://books.google.com/books?hl=en&lr=&id=FhPeRRKGLAUC&oi=fnd&pg=PR9&dq=Wolfram,+D.:+Applied+Informetrics+for+Information+Retrieval+Research.+Libraries+Unlimited+(2003)&ots=rUtg4fV_r9&sig=UgDKr1D9OQETbnyCt9fF6yRRELc)
- Wolfram, D. (2016). Bibliometrics, Information Retrieval and Natural Language Processing: Natural Synergies to Support Digital Library Research. In *Proceedings of the Joint Workshop on Bibliometric-enhanced Information*

- Retrieval and Natural Language Processing for Digital Libraries (BIRNDL2016)* (pp. 6–13). Retrieved from http://www.aclweb.org/old_anthology/W/W16/W16-1501.pdf
- Woodward, A. M. (1977). The roles of reviews in information transfer. *Journal of the American Society for Information Science*, 28(3), 175–180.
<https://doi.org/10.1002/asi.4630280306>
- Wright, M., Sumner, T., Moore, R., & Koch, T. (2007). Connecting digital libraries to eScience: the future of scientific scholarship. *International Journal on Digital Libraries*, 7(1), 1–4. <https://doi.org/10.1007/s00799-007-0030-9>
- Wu, F., & Huberman, B. A. (2004). Finding communities in linear time: a physics approach. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(2), 331–338. <https://doi.org/10.1140/epjb/e2004-00125-x>
- Wu, M., & Chen, S. (2014). Graduate students appreciate Google Scholar, but still find use for libraries. *The Electronic Library*, 32(3), 375–389.
<https://doi.org/10.1108/EL-08-2012-0102>
- Xiao, B., & Benbasat, I. (2007). E-commerce Product Recommendation Agents: Use, Characteristics, and Impact. *MIS Quarterly*, 31(1), 137–209. Retrieved from <http://aisel.aisnet.org/misq/vol31/iss1/9/>
- Xiao, Z., Che, F., Miao, E., & Lu, M. (2014). Increasing Serendipity of Recommender System with Ranking Topic Model. *Applied Mathematics & Information Sciences*, 8(4), 2041–2053. <https://doi.org/10.12785/amis/080463>
- Yang, Q., Zhang, S., & Feng, B. (2007). Research on Personalized Recommendation System of Scientific and Technological Periodical Based on Automatic Summarization. In *2007 First IEEE International Symposium on Information*

- Technologies and Applications in Education* (pp. 34–39). Ieee.
<https://doi.org/10.1109/ISITAE.2007.4409232>
- Yang, W., & Lin, Y.-R. (2013). A task-focused literature recommender system for digital libraries. *Online Information Review*, 37(4), 581–601.
<https://doi.org/10.1108/OIR-10-2011-0172>
- Yang, Z., & Davison, B. D. (2012). Venue Recommendation: Submitting your Paper with Style. *Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012, 1*, 681–686.
<https://doi.org/10.1109/ICMLA.2012.127>
- Yoo, I., & Mosa, A. S. M. (2015). Analysis of PubMed User Sessions Using a Full-Day PubMed Query Log: A Comparison of Experienced and Nonexperienced PubMed Users. *JMIR Medical Informatics*, 3(3).
<https://doi.org/10.2196/medinform.3740>
- Yoon, S.-H., Kim, S.-W., & Park, S. (2010). A link-based similarity measure for scientific literature. In *Proceedings of the 19th international conference on World wide web* (pp. 1213–1214). Raleigh, North Carolina, USA.
<https://doi.org/10.1145/1772690.1772880>
- Zarrinkalam, F., & Kahani, M. (2012). A New Metric for Measuring Relatedness of ScientificPapers Based on Non-Textual Features. *Intelligent Information Management*, 4(4), 99–107. <https://doi.org/10.4236/iim.2012.44016>
- Zarrinkalam, F., & Kahani, M. (2013). SemCiR: A citation recommendation system based on a novel semantic distance measure. *Program: Electronic Library and Information Systems*, 47(1), 92–112. <https://doi.org/10.1108/00330331311296320>
- Zhang, Z., & Li, L. (2010). A Research Paper Recommender System based on

Spreading Activation Model. In *The 2nd International Conference on Information Science and Engineering* (pp. 928–931). Ieee.

<https://doi.org/10.1109/ICISE.2010.5689417>

Zhu, H., & Zhou, M. (2008). Roles in Information Systems: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(3), 377–396. <https://doi.org/10.1109/TSMCC.2008.919168>

Zipf, G. (1949). Human Behavior and the Principle of Least Effort. In *Ed: Addison-Wesley*. Cambridge: MA: Addison-Wesley.

APPENDIX A: IRB APPROVAL DOCUMENT



Research Support Office

Reg. No. 200604393R

IRB-2014-09-017

27 October 2014

Professor Foo Shou Boon, Schubert
Wee Kim Wee School of Communication and Information

NTU INSTITUTIONAL REVIEW BOARD APPROVAL

Project Title: Examining the application of recommender systems for assisting researchers with their literature review and publication purposes

I refer to your application for ethics approval with respect to the above project.

The Board has deliberated on your application and noted from your application that your research involves collecting behavioral data from participants using questionnaires.

You have also confirmed that informed consent will be obtained from the participants and you have guaranteed the confidentiality of your participants' biodata obtained from them.

The documents reviewed are:

- a) NTU IRB application form dated **19 September 2014**
- b) Participant information sheet and consent form: version 1 dated **19 September 2014**
- c) Data collection form: version 1 dated **19 September 2014**

The Board is therefore satisfied with the bioethical consideration for the project and approves the ethics application under **Expedited** review. The approval period is from **27 October 2014** to **16 December 2015**. The NTU IRB reference number for this study is **IRB-2014-09-017**. Please use this reference number for all future correspondence.

The following protocol and compliances are to be observed upon NTU IRB approval

1. All research involving procedures greater than minimal risk on minors (individuals who are less than the legal age of 21 years old) requires IRB approved written Parental Consent and assent from the participant to be obtained before any research protocols can be administered. Minimal risk refers to an anticipated level of harm and discomfort that is no greater than that ordinarily encountered in daily life, or during the performance of routine educational, physical, or psychological examination.
2. Only the approved Participants Information Sheet and Consent Form should be used. It must be signed by each subject prior to initiation of any protocol procedures. In addition, each subject should be given a copy of the signed consent form.

Blk N2.1, B4-01, 76 Nanyang Drive, Singapore 637331 Nanyang Avenue, Singapore 639798
Tel : +65 6791 9857, Fax: 6793 2019
www.ntu.edu.sg



Research Support Office

3. Consent forms are important documents therefore they should be stored in the strictest arrangement. Loss of consent form would result in disciplinary action.
4. No deviation from, or changes of, the protocol should be initiated without prior written NTU IRB approval of an appropriate amendment.
5. The Principal Investigator should report promptly to NTU IRB regarding:
 - a. Deviation from, or changes to the protocol.
 - b. Changes increasing the risk to the subjects and/or affecting significantly the conduct of the trial
 - c. All serious adverse events (SAEs) which are both serious and unexpected.
 - d. New information that may affect adversely the safety of the subjects of the conduct of the trial.
 - e. Completion of the study.
6. Continuing Review Request/ Notice of Study completion form should be submitted to NTU IRB for the following:
 - a. Annual review: Status of the study should be reported to the NTU IRB at least annually using the Continuing Review Request/ Notice of Study completion form.
 - b. Study completion or termination: Continuing Review Request/ Notice of Study completion form is to be submitted within 4 to 6 weeks of study completion or termination.

A handwritten signature in black ink, appearing to read 'R. Lee'.

Professor Lee Sing Kong,
Chair, NTU Institutional Review Board
encl.

cc Chair, Wee Kim Wee School of Communication and Information
Members, NTU Institutional Review Board

APPENDIX B: STUDY I DETAILS

Part A: Participant Information and Consent Form

Study about Reviewers' and Authors' Experience in Citing Prior Literature in Manuscripts

Purpose of the research:

The purpose of this research is to conduct a study in order to gain insights on instances where researchers miss citing prior literature in their journal and conference paper manuscripts, from both reviewer and manuscript author perspectives. Additionally, the study also looks at three related areas - the information searching behavior of manuscript reviewers, manuscript authors' opinion on characteristics of research search tasks and manuscript authors' usage of paper recommendation features in academic databases and search engines. This survey is part of the research undertaken by Aravind Sesagiri Raamkumar, PhD, in the Wee Kim Wee School of Communication & Information (WKWSCI), NTU, Singapore. The Principal Investigator for this project is Dr. Schubert Foo from the Division of Information Studies, School of Communication & Information, Nanyang Technological University.

Eligibility:

The eligibility criteria for participation are 1.) You must be above 20 years of age, 2.) You must have published research articles in journals or conferences and 3.) You must have played the role of a reviewer for research articles in journals or conferences

What you will do in this study:

You will be provided with a questionnaire that contains the study related questions. It should not take more than 15-20 minutes to complete.

Risks:

There are no anticipated risks, beyond those encountered in daily life, associated with participating in this study.

Compensation:

You will receive **S\$10** upon completion of the full questionnaire.

Voluntary Withdrawal:

Your participation in this study is completely voluntary, and you may withdraw from the study at any time without penalty. Your decision to participate, decline, or withdraw participation will have no effect on your status or relationship with the Nanyang Technological University.

Confidentiality:

Your participation in this study will remain confidential, and your identity will not be stored with your data. Your responses will be assigned a code number that is not link to your name or other identifying information. All data and consent forms will be stored in a locked room. Results of this study may be presented at conferences and/or published in books, journals, and/or in the popular media.

Further Information:

If you have questions about the study or your rights as a participant in this study, please contact Dr. Schubert Foo, Division of Information Studies, Wee Kim Wee School of Communication & Information, Nanyang Technological University,
Email: SFOO@ntu.edu.sg

Who to contact about your rights in this study:

If you have any concerns about this study or your experience as a participant, you may contact the Institutional Review Board (IRB) at NTU at 65-65922495 (collect calls will be accepted if you state you are a study participant);email: irb@ntu.edu.sg

Agreement:

The purpose and nature of this research have been sufficiently explained and I understand that I am free to withdraw at any time without incurring any penalty. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. I agree to participate in this study.

☐ Yes

☐ No

Part B: Survey Questionnaire

Section 1: Demographics Details

- 1) Please provide your initials or name
- 2) Gender
[Male/Female]
- 3) Age
[18-24 years old/25-34 years old/35-44 years old/45-54 years old/55-64 years old/65 years or older]
- 4) Highest Education Level
[Post Graduate Diploma/Bachelor's Degree/ Master's Degree/Doctor's Degree/Others]
- 5) Current Position
[Graduate Research Student/Research Associate/Research Fellow/Lecturer/Senior Lecturer/Assistant Professor/Associate Professor/Professor/Others]
- 6) Parent School
- 7) Primary Discipline
- 8) Have you played the role of a reviewer for journals or conferences in the past?
[Yes/No]
- 9) Which of the below research experience level describes you the best?
[Expert/Advanced/Intermediate/Beginner]

Section 2: Reviewers' Experience with Citation of Prior Literature in Manuscripts during Manuscripts Review

- 10) For how many years, have you been reviewing journal and conference papers?
- 11) While reviewing journal paper manuscripts, I have come across instances where...

[Never/Rarely/Sometimes/Very Often/Always]

- a.) Authors have missed citing important/seminal papers in their manuscripts
- b.) Authors have missed citing topically relevant papers in their manuscripts
- c.) There was need for authors to cite more papers in the Literature Review/Related Work sections of their manuscripts
- d.) Authors have cited irrelevant papers in their manuscripts

12) While reviewing conference paper manuscripts, I have come across instances where...

[Never/Rarely/Sometimes/Very Often/Always]

- a.) Authors have missed citing important/seminal papers in their manuscripts
- b.) Authors have missed citing topically relevant papers in their manuscripts
- c.) There was need for authors to cite more papers in the Literature Review/Related Work sections of their manuscripts
- d.) Authors have cited irrelevant papers in their manuscripts

13) Please provide your opinion on factors affecting authors' citing behavior of prior literature in manuscripts

[Strongly Disagree/Disagree/Neutral/Agree/Strongly Agree]

- a) Authors miss citing relevant papers particularly when the research topic is interdisciplinary in nature
- b) Authors' tendency of not citing the expected papers in their manuscripts, can be attributed to the lack of overall research experience
- c) Authors' tendency of not citing the expected papers in their manuscripts, can be attributed to the lack of experience in the particular research area

- d) Authors' inability in citing the relevant literature is one of the key reasons for sending the manuscript back for revision
- e) Authors' inability in citing the relevant literature is one of the key reasons for rejecting the manuscript

Section 3: Authors' Experience with Citation of Prior Literature in Manuscripts during Manuscripts Review

14) For how many years, have you been writing papers for journals and conferences?

15) During submission of journal paper manuscripts, I have come across instances where...

[Never/Rarely/Sometimes/Very Often/Always]

- a.) Reviewers have indicated that I have not cited certain important/seminal papers in the manuscript
- b.) Reviewers have indicated that I have not cited certain topically related paper(s), in the manuscript
- c.) Reviewers have indicated that I have not cited enough papers in the Literature Review section of the manuscript
- d.) Reviewers have indicated that I have cited irrelevant papers, in the manuscript

16) During submission of conference paper manuscripts, I have come across instances where...

[Never/Rarely/Sometimes/Very Often/Always]

- a.) Reviewers have indicated that I have not cited certain important/seminal papers in the manuscript
- b.) Reviewers have indicated that I have not cited certain topically related paper(s), in the manuscript

- c.) Reviewers have indicated that I have not cited enough papers in the Literature Review section of the manuscript
- d.) Reviewers have indicated that I have cited irrelevant papers, in the manuscript

Section 4: Researchers' Tasks in Literature Review and Manuscript Writing Process

17) How often do you feel the need for assistance in...

[Never/Rarely/Sometimes/Very Often/Always]

- a.) Identifying seminal/important papers that are to be read as a part of the literature review in your research study
- b.) Identifying papers that are topically similar to the papers that you have already read as part of your literature review
- c.) Identifying papers related to your research, from disciplines other than your primary discipline
- d.) Identifying papers for particular placeholders in your manuscript
- e.) Identifying papers that must be necessarily cited in your manuscripts

Section 5: Usage of Academic Information Sources

18) I am aware of the paper recommendation feature in online academic systems such as academic databases, search engines and digital libraries

[Strongly Disagree/Disagree/Neutral/Agree/Strongly Agree]

19) I have used the below stand-alone recommender systems during literature search sessions

[Refseer/Theadvisor/CSSeer/PubChase/Others/Not used any of the above systems/Haven't heard of any of the above systems]

20) I have used the academic search engine 'Google Scholar' during my information search sessions

[Never/Rarely/Sometimes/Very Often/Always]

21) I have used the 'Related Articles' feature in Google Scholar

[Never/Rarely/Sometimes/Very Often/Always]

22) I have used the academic database 'Web of Science' during my information search sessions

[Never/Rarely/Sometimes/Very Often/Always]

23) I have used the 'View Related Records' feature in Web of Science

[Never/Rarely/Sometimes/Very Often/Always]

24) I have used the academic database 'Scopus' during my information search sessions

[Never/Rarely/Sometimes/Very Often/Always]

25) I have used the 'Related Documents' feature in Scopus

[Never/Rarely/Sometimes/Very Often/Always]

26) I have used the academic database 'IEEE Xplore' during my information search sessions

[Never/Rarely/Sometimes/Very Often/Always]

27) I have used the 'Similar' feature in IEEE Xplore

[Never/Rarely/Sometimes/Very Often/Always]

28) I have used the academic database 'Science Direct' during my information search sessions

[Never/Rarely/Sometimes/Very Often/Always]

29) I have used the 'Recommended Articles' feature in Science Direct

[Never/Rarely/Sometimes/Very Often/Always]

30) I have used the academic database 'Springer Link' during my information search sessions

[Never/Rarely/Sometimes/Very Often/Always]

31) I have used the 'Related Content' feature in Springer Link

[Never/Rarely/Sometimes/Very Often/Always]

32) I have used the academic search engines 'PubMed' during my information search sessions

[Never/Rarely/Sometimes/Very Often/Always]

33) I have used the 'Related Citations' feature in PubMed

[Never/Rarely/Sometimes/Very Often/Always]

APPENDIX C: PERFORMANCE TWEAKS FOR REC4LRW SYSTEM

During the course of the development and pilot testing of the Rec4LRW system, certain performance tweaks were applied to improve the retrieval time of the recommendations at both task and system level. These tweaks are described as follows.

Generic Tweaks

Assignment of Information Cue Labels

The four information cue labels *Popular*, *Recent*, *Survey/Review* and *High Reach* (described in Section 4.4.3.1 in Chapter 4), were displayed alongside the recommendations for all the three tasks. The label assignment of *Recent* and *Survey/Review* cues was a simple process and it was performed dynamically whenever the task was run. *Recent* papers were identified by the Publication Date metadata field. If a paper's publication year ranges from 2009 to 2011, it was identified a recent paper as the ACM DL corpus comprised of papers published until 2011. The *Survey/Review* papers identification was performed by searching for the keywords 'review' or 'literature survey' in the Article Title and Abstract metadata fields of the paper. Hence, these two cues did not cause any performance issue in the system

On the other hand, *Popular* and *High Reach* cues did cause delay in the label assignment as these two cues deal with aggregated data. The former was applicable for a paper with citation count in the top 5% percentile of the citation counts for the parent topic while the latter was applicable for papers with references count in in the top 5% percentile of the references counts for the parent topic. The label assignment for these

two cues was initially done on-the-fly. This method caused a delay of about five seconds in the overall retrieval time. To resolve this issue, the lower threshold value for these two cue labels was pre-computed for the 43 research topics used in the user evaluation study and the values were stored as two columns in a MySQL table. After this tweak was applied, the front-end PHP script just had to check whether a particular paper's reference and citation counts were equal or higher than the threshold values so that the labels could be assigned for the particular paper. This tweak helped in reducing the retrieval time by about five seconds for each of the three tasks.

Periodic Clearance of Files in the Rec4LRW Server

As mentioned earlier in Section 4.4.1 in Chapter 4, the technology stack of Rec4LRW system involved the usage of both PHP and JAVA. The inter-connection between PHP and JAVA was established through BATCH files and TEXT files. While executing a recommendation task, PHP script executes a BATCH file that in-turn executes a JAVA command to trigger a corresponding program. The JAVA program interacted with the MySQL database. The data retrieved from the database was then stored in TEXT files. The data from these TEXT files are in-turn picked up the PHP scripts for display in the web pages. Whenever a recommendation task was run, a new TEXT file for the particular username and timestamp was created and stored in the server. As the recommendation tasks were run for multiple times over a period of few days, around 50 TEXT files were created every day in the server and the presence of these files seemed to affect the overall retrieval speed of the tasks. Hence, as a resolution measure, these TEXT files were deleted on a daily basis to clear the server space. This helped in improving the performance to about two seconds for each task.

Task-level Tweaks

Task 1 is not a very computationally intensive task as the retrieval process is similar to basic keyword-based information retrieval (IR) task. After the application of the generic tweaks, this task took around four seconds to retrieve and display the recommendations in the task screen. During peak hours when there were multiple parallel users using the Rec4LRW system, the retrieval time increased to about eight seconds. Since, the retrieval time was on an acceptable level, there were no task-specific tweaks applied for this task.

Task 2 was the most computationally intensive task in the Rec4LRW system since it had two modules for discovering similar papers (refer Section 4.3.4.2 in Chapter 4). Out of these two modules, the topical similarity module invoked an Information Retrieval (IR) process while the chaining similarity module invoked collaborative filtering (CF) process and database (DB) based filtering process. The IR and DB processes are executed using Apache Lucene and MySQL respectively while the CF process is executed using Apache Mahout. The initial task execution time was about 55 to 60 seconds. The IR and DB processes offered very little scope for performance improvements. However, there was scope for pre-computed values in the CF process. As mentioned in Section 4.3.4.2, the item-based CF variant (IBCF) was used. The IBCF algorithm is memory-based since the similarity is calculated between the items in the user-item matrix, unlike the UBCF variant. In Apache Mahout, there is a provision to compute the similarity between items on an offline basis. These pre-computed values could be subsequently added to the CF algorithm so that the identification of recommendations could be speeded up. Using this approach, the CF process was improved to around 50 seconds. All the other generic tweaks, helped in improving the speed to about 45 seconds. However, many participants in the user

study felt that the task run time was still high. Similar to Task 1, Task 3 was not a computationally intensive task. The whole retrieval process took about 6-10 seconds to complete, there was not much incentive to further reduce the retrieval time. Secondly, there were no complaints from the user study participants regarding this task's performance.

APPENDIX D: CASE STUDY FOR GREY LITERATURE BOOSTING TECHNIQUE

Part A: Analysis of Bibliographic References in ACM DL Extract

The ACM DL dataset (refer Section 5.2 of Chapter 5) was used for this analysis. The percentages of grey literature GL and non-GL references by different article-types are shown in Figure A.1. The GL Percentage (GLP) (refer Section 4.2.2 of Chapter 4) on the whole for proceedings and periodicals was 17.61% and 14.48%, indicating the nature of the publications. In ACM DL, journals, magazines and transactions constitute periodicals. Under periodicals, articles are mainly classified as research articles while some journals also allow review/survey articles. Survey articles had the lowest GLP (12.86%). Contrastingly, referencing behavior in conferences, symposiums and workshops was observed to be comparatively less stringent. There are different article-types under proceedings. The types demos (18.71%) and tutorials (17.98%) had the highest GLP as the references in articles of such types are of miscellaneous nature. The current examination is limited to quantifying the referencing of GL articles in the ACMD DL extract. In-depth qualitative studies need to be conducted to identify the citation motivations of authors based on the citation contexts of GL references in articles.

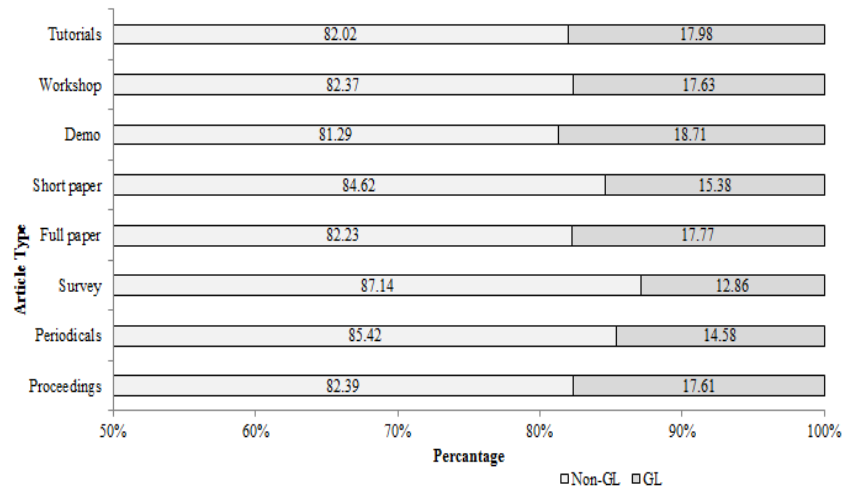


Figure A 1. GL vs Non-GL references percentage in ACM DL articles

Part B: Experimentation with GL Boosting Technique

In academic search systems and digital libraries, the article metadata along with full text are generally indexed. Papers are retrieved based on the matching of search keywords and the indexed text. The final results are sorted using citation count, search keyword relevance, recency, and other factors. Bibliographic references from articles are also indexed in some of these systems although these items are rarely displayed in the search results. For promoting GL references in IR systems, the technique of document boosting can be adopted. A separate field known as ‘boosting weight’ can be added to each document (research paper) when the indexing process is performed. The boosting weights can be used during the retrieval process to compute new similarity score on top of the basic similarity score. By using this technique, GL materials that get additional weightage due the boosting rules will have a higher probability of getting a better rank in the search results, thereby increasing their visibility to users.

The results of an experiment carried out to validate the GL references boosting technique applicable for scientific paper IR systems, are briefly discussed. For this

experiment, the proceedings articles of the ACM DL dataset were used. A total of 103,739 articles and corresponding 2,320,345 references were indexed in a Lucene index file. The fields indexed were *article id*, *article title*, *article abstract*, *article-type* and the *boosting weight* (gl_w). For all the proper full-text articles, article-type was set as ‘inproceedings’ and for the references, the type was set based on the reference-type identified using the AnyStyle reference parsing service. Okapi BM25(K Sparck Jones et al., 2000) ($k=2$, $f=0.75$) was used for the initial similarity score calculation. The boosting weight rules used are provided as follows.

- *Rule 1: If the article or the reference is of non-GL type, the boosting weight is 1.0*
- *Rule 2: If the reference is of GL type and its reference count is more than 2, the boosting weight is 1.5,*
- *Rule 3: If the reference type is a thesis, the boosting weight is 1.25.*

Rules 2 and 3 are the two scenarios used for boosting GL materials in this case study. The boosting weights for these rules are samples for this study since the aim is elevate certain GL articles. A separate experiment is required to arrive at ideal weights for boosting rules.

Ten research topics were used as the search keywords for the experiment. These topics are part of the top 50 author specified keywords in the ACM DL. For each research topic, a search was performed and the top 100 results were retrieved based on the BM25 similarity score. For the BM25 similarity score calculation, the title and abstract fields of the articles were used. From the retrieved 100 results, the top 20 results were ranked using four techniques. The proposed GL Boosting Technique (GBT) was benchmarked against two traditional ranking techniques and one combined

ranking technique. The three benchmarking ranking techniques were Citation Count Technique (CCT) in which the results were ranked based on article's citation count, BM25 Similarity Score Technique (BST) in which the ranking was based on the computed BM25 similarity score and thirdly, the Combined Score Technique (CST) where the values of citation count and BM25 similarity score were added to form a combined score. In the proposed GBT, the ranking was based on the values which were computed by multiplying the BM25 similarity score with the boosting weight (gl_w).

In order to perform a proper evaluation, a novel evaluation metric Grey Literature Availability Measure (GLAM) had to be conceptualized that accounted for both the retrieved GL articles count as well as the corresponding ranks. GLAM is based on two base metrics GL Count (GLC) which is the count of GL materials that are retrieved in the query and in-query MRR (iMRR), a modified version of Mean Reciprocal Rank (Voorhees, 1999). Unlike MRR which is calculated across a set of queries, iMRR is calculated within a single query. First, the reciprocal ranks of GL articles in the query are identified. Second, the sum of the reciprocal ranks is divided by the GL articles count (GLC) to form the iMRR value. The main evaluation metric Grey Literature Availability Measure (GLAM) is calculated by adding up GLC and iMRR for each query. Higher GLAM values indicate higher presence of GL articles in the query results, along with better ranks.

The evaluation was done with two ranked lists (N@10 and N@20) for each research topic. Figures A.2 and A.3 provide the GLAM values computed using the four techniques for the 10 research topics at N@10 and N@20 respectively. For N@10, GL boosted technique (GBT) has higher GLAM than the benchmarking techniques as the GL count (GLC) is higher for all research topics. For N@20, except

for two research topics ‘interaction design’ and ‘wireless networks’ where CCT technique has a higher GLAM, the GBT technique produces the best results for all other research topics. Thus, the experiment shows the GL boosting technique produces the expected results for majority of the input research topics.

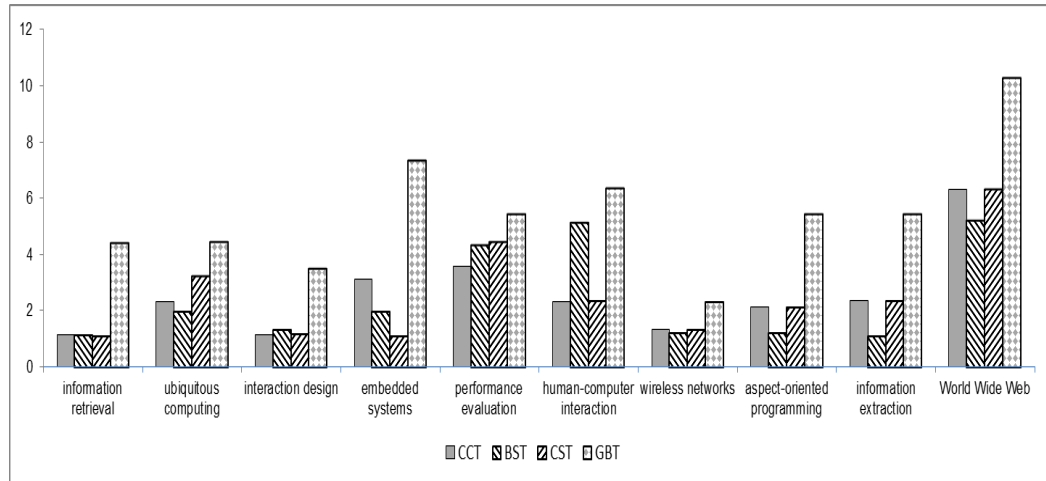


Figure A.2. Comparison of GLAM values for the four techniques (N@10)

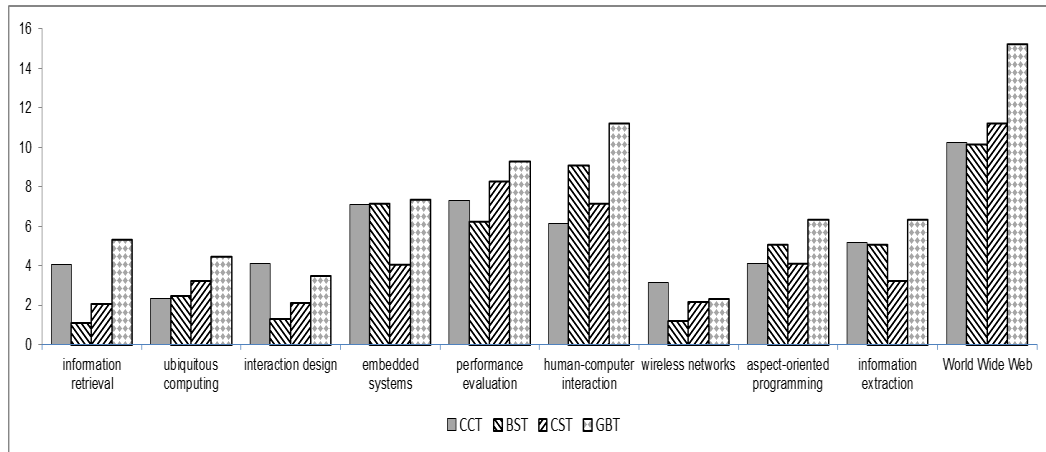


Figure A.3. Comparison of GLAM values for the four techniques (N@20)

APPENDIX E: CASE STUDY WITH THE COVERAGE FEATURES

Experiment Details

In order to analyze the performance of the coverage techniques *TPC* and *TC* (refer Section 5.2.3 in Chapter 5) at a detailed level, an experiment was conducted with a restricted set of papers which are about a single broad topic. “Information Retrieval (IR)” was chosen as the research topic for this study as it is a broad area with 71 sub-topics in the ACM 2012 taxonomy (ACM, 2015). From a base extract of the ACM DL dataset (refer Section 5.2 in Chapter 5) covering papers published between 1951 and 2010, 1473 IR papers were initially shortlisted. 976 papers were finally chosen as the sample set as these papers had complete information in fields such as keywords, abstract, bibliography and full text. 21,243 references and citations data of the sample set papers were used to build the references and citations network.

TPC and *TC* were benchmarked with HITS (Kleinberg, 1999) value of a paper, referred to as *HC* in this experiment. The coverage values for the two techniques were measured for all the sample set papers and the top 20 papers were shortlisted for comparison purposes since most users tend to select resources within the top 20 results that are displayed within the two pages of the search results (Van Deursen & Van Dijk, 2009). The four requirements for the reading list (refer Section 4.3.3.1 in Chapter 4) were the four evaluation criteria used for this experiment. The measurement for the evaluation criteria was the same as used in the offline experiment of the AKR technique (refer Section 5.3 of Chapter 5).

Results and Discussion

The experiment results for the four evaluation criteria are provided in Table A.1 for the three techniques. The top 20 papers from the three techniques are displayed in Table A.2, A.3 and A.4 respectively where the columns Ref. Ct and Cit. Ct refer to references count and citations count of the papers. In these three tables, survey papers and also recent papers published between 2008 and 2010 are highlighted in italics.

Table A.1

Criteria Counts for the Three Techniques

Criteria Count	HC	TPC	TC
Number of Edges in Subgraph (Diversity)	14	5	18
Number of Literature Survey Papers	6	1	3
Number of Papers with Citation Count above 100	9	16	16
Number of Recent Papers	3	7	5

Table A.2

Top 20 papers from HC technique

Rank	Title (year)	Ref. Ct	Cit. Ct
1	<i>Inverted files for text search engines (2006)</i>	205	900
2	<i>Information retrieval on the web (2000)</i>	231	735
3	<i>Information storage and retrieval: a survey and functional description (1977)</i>	238	19
4	<i>Web mining research: a survey (2000)</i>	128	174

Rank	Title (year)	Ref.	Cit.
		Ct	Ct
			2
5	Information science in a Ph.D. computer science program (1969)	200	9
6	Information retrieval on the semantic web (2002)	28	225
7	Building efficient and effective metasearch engines (2002)	83	413
8	Refinement of TF-IDF schemes for web pages using their hyperlinked neighboring pages (2003)	26	68
9	<i>A survey of Web clustering engines (2009)</i>	113	277
10	Collection synthesis (2002)	47	64
11	Using information scent to model user information needs and actions and the Web (2001)	26	401
12	Social network document ranking (2010)	36	20
13	An indexing model of HTML documents (2003)	25	12
14	A Hybrid Technique for English-Chinese Cross Language Information Retrieval (2008)	49	20
15	Efficient on-line index maintenance for dynamic text collections by using dynamic balancing tree (2007)	15	28
16	Superimposing codes representing hierarchical information in web directories (2001)	16	10
17	Information filtering and information retrieval: two sides of the same coin? (1992)	33	166 4
18	<i>Is this document relevant? probably?: a survey of probabilistic models in information retrieval (1998)</i>	90	270
19	Microsearch: A search engine for embedded devices used in	36	21

Rank	Title (year)	Ref. Ct	Cit. Ct
pervasive computing (2010)			
20	Merging techniques for performing data fusion on the web (2001)	27	30

Table A. 3

Top 20 papers from TPC technique

Rank	Title (year)	Ref. Ct	Cit. Ct
1	Information retrieval on the semantic web (2002)	28	225
2	Methods and metrics for cold-start recommendations (2002)	31	102 4
3	Stuff I've seen:a system for personal information retrieval and re-use (2003)	32	813
4	Designing a digital library for young children (2001)	20	156
5	<i>Information retrieval on the web</i> (2000)	231	735
6	A cluster-based resampling method for pseudo-relevance feedback (2008)	35	142
7	A case for interaction: a study of interactive information retrieval behavior and effectiveness (1996)	13	398
8	Query dependent pseudo-relevance feedback based on wikipedia (2009)	34	135
9	Scatter/gather browsing communicates the topic structure of a very large text collection (1996)	11	245

Rank	Title (year)	Ref.	Cit.
		Ct	Ct
10	Answer Garden 2: merging organizational memory with collaborative help (1996)	25	399
11	Using information scent to model user information needs and actions and the Web (2001)	26	401
12	A knowledge-based search engine powered by wikipedia (2007)	21	153
13	Evaluation over thousands of queries (2008)	21	79
14	Sources of evidence for vertical selection (2009)	20	126
15	Interactive textbook and interactive Venn diagram: natural and intuitive interfaces on augmented desk system (2000)	21	117
16	Categorizing web queries according to geographical locality (2003)	21	171
17	Automatic query generation for patent search (2009)	12	67
18	Probabilistic query expansion using query logs (2002)	18	415
19	Extending average precision to graded relevance judgments (2010)	27	51
20	Learning in a pairwise term-term proximity framework for information retrieval (2009)	19	47

Table A.4

Top 20 papers from TC technique

Rank	Title (year)	Ref	Cit.
		. Ct	Ct

Rank	Title (year)	Ref	Cit.
		. Ct	Ct
1	A vector space model for automatic indexing (1975)	7	57 80
2	<i>Information retrieval on the web</i> (2000)	231	73 5
3	Information retrieval on the semantic web (2002)	28	22 5
4	Cross-language information retrieval based on parallel texts and automatic mining of parallel texts from the Web (1999)	19	32 0
5	A Hybrid Technique for English-Chinese Cross Language Information Retrieval (2008)	49	20
6	What the query told the link: the integration of hypertext and information retrieval (1997)	37	11 7
7	Geospatial mapping and navigation of the web (2001)	22	28 2
8	Stuff I've seen: a system for personal information retrieval and re-use (2003)	32	81 3
9	<i>Is this document relevant? Probably? :a survey of probabilistic models in information retrieval</i> (1998)	90	27 0
10	<i>Inverted files for text search engines</i> (2006)	205	90 0
11	A case for interaction: a study of interactive information retrieval behavior and effectiveness (1996)	13	39 8
12	Biterm language models for document retrieval (2002)	10	10

Rank	Title (year)	Ref . Ct	Cit. Ct
			7
13	Discovering key concepts in verbose queries (2008)	34	19 7
14	Statistical transliteration for english-arabic cross language information retrieval (2003)	22	15 6
15	Categorizing web queries according to geographical locality (2003)	21	17 1
16	Query term disambiguation for Web cross-language information retrieval using a search engine (2000)	19	60
17	Modeling and visualizing geo-sensitive queries based on user clicks (2008)	14	16
18	Integration of news content into web results (2009)	34	10 5
19	An improved markov random field model for supporting verbose queries (2009)	32	47
20	Extended Boolean information retrieval (1983)	28	10 48

For the diverse papers inclusivity criterion, the least number of edges was for the subgraph formed with the top 20 papers from the *TPC* technique ($n=5$) (refer Table A.1) while the highest was for *TC* technique ($n=18$). The inference is that *TPC* technique is better at providing a diverse set of papers about different sub-topics in IR. This result is validated by the inclusion of non-IR keywords in the base set formation for *TPC* technique, thereby increasing the scope for a heterogeneous mix of papers. On

the other hand, *TC* and *HC* subgraphs had more connections as the top 20 papers are of IR-exclusive nature. Literature survey/review papers inclusion was clearly the highest for the *HC* technique ($n=6$). The finding vindicates the working mechanism of HITS algorithm as it is expected to find hub nodes which point to many authority nodes. The references count is evidently higher for survey papers as these papers review the state-of-the-art in the particular research area. The number of survey papers in the top 20 papers of *TC* ($n=3$) and *TPC* ($n=1$) was comparatively lower as these techniques do not necessarily give exclusive importance to papers with high reference counts, at the conceptual level. The survey paper titles are highlighted in italics in the Tables A.2, A.3 and A.4.

For the popular papers inclusivity criterion, both *TPC* and *TC* techniques were good at providing high coverage values to highly cited papers ($n=16$) while interestingly for *HC* technique, the count was lower ($n=9$) even though the technique identified more number of survey papers. This is due to the tendency of the *HC* technique in giving higher coverage scores to papers with high reference counts. In the case of recent papers inclusivity, the least number of recent articles ($n=3$) was for *HC* technique while the count was better for *TPC* ($n=7$) and *TC* ($n=5$). None of the techniques have temporal preferences in the conceptual model, therefore the results for this criterion need to be validated with papers for different research topics. It is assumed that if a recent paper has a higher references count, it is expected to get good coverage scores with the *HC* technique. In the current experiment, only one recent survey paper ‘A survey of Web clustering engines’ published in 2009 received higher coverage score (rank 9 in *HC* top 20 papers).

Table A.5

Ranks of the three techniques

Rank	HC	TPC	TC
Number of Edges in Subgraph (Diversity)	2	1	3
Number of Literature Survey Papers	1	3	2
Number of Papers with Citation Count above 100	3	1	1
Number of Recent Papers	3	1	2

Based on the performance of the three techniques with the four evaluation criteria, ranks have been assigned in Table A.5. *TPC* technique performs the best for diverse, popular and recent papers while *HC* technique performs the best for survey papers. The coverage metric can be implemented in citations databases, digital libraries and academic search engines for ranking research papers. The most specific usage of this metric is in ranking papers towards building a reading list of papers for literature review. Even though, it could be argued that the reading list generation technique should be standardized for all researchers, special preferences could also be given. For instance, when recommending research papers to novice researchers, a substantial amount of survey articles could be recommended. Alternatively for experienced researchers, a diverse set of papers about different sub-topics of a particular research area could be recommended. Therefore, the *TPC* technique and *HC* technique could be employed as per the specific needs of the users, for providing better results. Alternatively, they could be combined to form a single score. This alternative approach was attempted in the offline evaluation experiment (refer Section 5.3 of Chapter 5) and the assumption proved to be correct as the integrated approach

seemed to produce the best results, satisfying most of the requirements of the reading list.

**APPENDIX F: FACETED CLASSIFICATION SCHEME FOR A
TASK FROM LI & BELKIN (2008)**

Category	Facets	Sub-facets	Values	Operational definitions
Generic facet of task	Source of task		Internal generated	A task motivated by a task doer. It is a self-motivated task
			Collaboration	A task motivated through discussion of a group of people
			External assigned	A task assigned by task setters based on their individual purpose
	Task doer		Individual	A task conducted by one task doer
			Individual in a group	A task assigned and completed by different group members separately, though they are in a group
			Group	A task conducted by a group of people (at least two people)
	Time	Frequency	Unique	A task conducted at the first time
			Intermittent	A task conducted more than one time but assessed by task doer as not frequently conducted
			Routine	A task assessed by task doer as frequently conducted
		Length	Short-term	A task which could be finished within a short time period (e.g. less than one month)
			Long-term	A task which has to be finished within a long time period (e.g. more than one month)
		Stage	Beginning	A task which just launched
			Middle	A task which has been running for a while and in the middle way
			Final	A task which is almost done or has been

Category	Facets	Sub-facets	Values	Operational definitions
				completed
	Product		Physical (for WT)	A task which produces a physical product
			Intellectual (for WT and ST)	A task which produces new ideas or findings
			Decision/Solution (for WT)	A task which involves decision making or problem solving
			Factual information (for ST)	A task locating facts, data, or other similar information items in information systems
			Image (for ST)	A task locating images in information systems
			Mix product (for ST)	A task locating different types of information items in information systems
	Process		One-time task	A task accomplished through one process without repeated procedures
			Multi-time task	A task accomplished through repeatedly engaging in the same or similar process
	Goal	Quality	Specific goal	A task with explicit or concrete goals
			Amorphous goal	A task with abstract goals
			Mixed goal	A task with both concrete and abstract goals
		Quantity	Multi-goal	A task with two or more goals
			Single-goal	A task with only one goal
Common attributes of task	Task characteristics	Objective task complexity	High complexity	A task which involves significantly more paths during engaging in the task
			Moderate	A task which may involve a few paths but not significantly more during engaging in the task
			Low complexity	A task which involves a single path during engaging in the task

Category	Facets	Sub-facets	Values	Operational definitions
		Interdependence	High interdependence	A task conducted through collaboration of a group of people (at least two people)
			Moderate	A task conducted by one task doer with suggestions or help from other people or group members
			Low interdependence	A task conducted by one task doer without any help from other people
	User's perception of task	Salience of a task	High salience	A task assessed by the task doer as highly important
			Moderate	A task assessed by a task doer as moderate importance or the degree of salience of the task depends on specific situations
			Low salience	A task assessed by the task doer as unimportant
		Urgency	Immediate (urgent)	A task assessed by a task doer as highly urgent
			Moderate	A task assessed by the task doer as moderately urgent or the degree of urgency of the task depends on specific situations
			Delayed (not urgent)	A task assessed by the task doer as no urgency
		Difficulty	High difficulty	A task assessed by a task doer as high difficulty
			Moderate	A task assessed by a task doer as moderate difficulty or the degree of difficulty of the task depends on specific situations
			Low difficulty	A task assessed by a task doer as no difficulty or easy to complete
		Subjective	High complexity	A task assessed by a task doer as highly

Category	Facets	Sub-facets	Values	Operational definitions
		task		complex
		complexity	Moderate	A task assessed by a task doer as moderately complex or the degree of complexity of the task depends on specific situations
			Low complexity	A task assessed by a task doer as simple
		Knowledge of task topic	High knowledge	A task assessed by a task doer as highly knowledgeable on the task-related topic
			Moderate	A task assessed by a task doer as moderately knowledgeable on the task-related topic or the degree of knowledge on the task topic depends on specific situations
			Low knowledge	A task assessed by a task doer as unknowledgeable on the task-related topic
		Knowledge of task procedure	High knowledge	A task assessed by a task doer as highly knowledgeable on the method or procedures for completing the task
			Moderate	A task assessed by a task doer as moderately knowledgeable on the method or procedures to completing the task or the degree of knowledge on the method or procedures depends on specific situations
			Low knowledge	A task assessed by the task doer as not knowledgeable on the method or procedures for completing the task

APPENDIX G: LIST OF 186 TOPICS USED FOR OFFLINE EVALUATION EXPERIMENT IN STUDY II

Topic
abstract interpretation
access control
active learning
ad hoc networks
affective computing
anomaly detection
approximation algorithms
artificial intelligence
aspect-oriented programming
assistive technology
augmented reality
autonomic computing
cloud computing
code generation
collaborative filtering
collaborative learning
combinatorial optimization
computational complexity
computer games
computer graphics
computer science education
computer vision
computer-mediated communication
congestion control
context awareness
data integration
data mining
data streams

Topic
data structures
design patterns
digital libraries
direct manipulation
distributed computing
distributed systems
domain-specific languages
dynamic analysis
dynamic programming
electronic commerce
embedded systems
empirical study
energy efficiency
evolutionary algorithms
evolutionary computation
eye tracking
fault tolerance
feature extraction
feature selection
formal methods
formal verification
functional programming
game design
game theory
garbage collection
genetic algorithms
genetic programming
global illumination
graph theory
grid computing
human computer interaction

Topic
human factors
human-computer interaction
human-robot interaction
image processing
image retrieval
information extraction
information retrieval
information security
information systems
information technology
information visualization
input devices
interaction design
interaction techniques
interface design
intrusion detection
knowledge management
linear programming
link analysis
load balancing
local search
location-based services
logic programming
low power
machine learning
mechanism design
memory management
mixed reality
mobile ad hoc networks
mobile computing
mobile devices

Topic
mobile phones
model checking
motion capture
multi-agent systems
multi-objective optimization
multiagent systems
multimodal interaction
natural language processing
network security
neural networks
non-photorealistic rendering
object-oriented programming
open source
operating systems
parallel computing
parallel processing
parallel programming
participatory design
particle swarm optimization
pattern matching
pattern recognition
performance analysis
performance evaluation
performance modeling
pervasive computing
power management
program analysis
program transformation
program verification
programming languages
project management

Topic
quality of service
query expansion
query optimization
query processing
question answering
real-time systems
recommender systems
reinforcement learning
relevance feedback
requirements engineering
resource allocation
reverse engineering
search engines
semantic web
sensor networks
social computing
social interaction
social media
social network analysis
social networks
software architecture
software development
software engineering
software evolution
software maintenance
software product lines
software quality
software reuse
software testing
speech recognition
static analysis

Topic
support vector machines
system design
tangible interfaces
tangible user interfaces
text classification
text mining
transactional memory
type inference
type systems
ubiquitous computing
usability evaluation
usability testing
user experience
user interface design
user interfaces
user modeling
user studies
user-centered design
video retrieval
virtual environments
virtual machines
virtual reality
virtual worlds
volume rendering
wearable computing
web 2
web accessibility
web applications
web mining
web search
web services

Topic
wireless networks
wireless sensor networks
world wide web

APPENDIX H: STUDY II DETAILS

Part A: Participant Information and Consent Form

User Evaluation of the Rec4LRW System

Purpose of the research:

The purpose of this research is to conduct a user evaluation study of the Rec4LRW system which is a scientific paper recommender system built for assisting researchers in literature review and manuscript writing. This system has been built specifically for three tasks – (1) Building an initial reading list of research papers, (2) Finding similar papers based on a set of papers and (3) Shortlisting papers from reading list for inclusion in manuscript. This research is part of the research undertaken by Aravind Sesagiri Raamkumar, PhD student, in the Wee Kim Wee School of Communication & Information (WKWSCI), Nanyang Technological University (NTU), Singapore. The Principal Investigator for this project is Dr. Schubert Foo from the Division of Information Studies, WKWSCI, NTU.

Eligibility:

The eligibility criteria for participation are 1.) You must be above 20 years of age and 2.) You must have published research paper(s) in journals or conferences

What you will do in this study:

You will have to execute the three tasks of the Rec4LRW system by following the provided instructions. After completing each task, you will be provided with a questionnaire that contains the evaluation related questions. The whole study should not take more than 45 minutes to complete.

Risks:

There are no anticipated risks, beyond those encountered in daily life, associated with participating in this study.

Compensation:

You will receive **S\$20** upon completion of the full experiment.

Voluntary Withdrawal:

Your participation in this study is completely voluntary, and you may withdraw from the study at any time without penalty. Your decision to participate, decline, or withdraw participation will have no effect on your status or relationship with the Nanyang Technological University.

Confidentiality:

Your participation in this study will remain confidential, and your identity will not be stored with your data. Your responses will be assigned a code number that is not link to your name or other identifying information. All data and consent forms will be stored in a locked room. Results of this study may be presented at conferences and/or published in books, journals, and/or in the popular media.

Further Information:

If you have questions about the study or your rights as a participant in this study, please contact Dr. Schubert Foo, Division of Information Studies, Wee Kim Wee School of Communication&Information, Nanyang Technological University,

Email: SFOO@ntu.edu.sg

Who to contact about your rights in this study:

If you have any concerns about this study or your experience as a participant, you may contact the Institutional Review Board (IRB) at NTU at 65-65922495 (collect calls will be accepted if you state you are a study participant);email: irb@ntu.edu.sg

Agreement:

The purpose and nature of this research have been sufficiently explained and I understand that I am free to withdraw at any time without incurring any penalty. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. I agree to participate in this study.

☐ Yes

☐ No

Part B: Rec4LRW Experiment User Guide

Introduction to the Rec4LRW System

Three Tasks in the Rec4LRW System

The Rec4LRW (Scientific Paper Recommender System for Literature Review and Writing) system is a task-based scholarly system meant for assisting researchers in finding research papers for their literature review and also for shortlisting papers during the manuscript writing stage. It is particularly aimed at helping beginners who are new to academic research and other researchers who are venturing into new research areas that are different from their areas of expertise. There are three tasks offered by the system.

- Task 1: Building an initial reading list of research papers for literature review
- Task 2: Finding similar papers based on a seed set of papers
- Task 3: Shortlisting papers from reading list for inclusion in manuscript

Task 1 is typically performed at the start of the literature review where the intent is to find a set of papers that will help the researchers in gaining a decent overview of the research topic. The process of building a reading list is heuristic-based in principle. Earlier systems have taken the approach of mainly retrieving seminal papers for building this list. However, the Rec4LRW system identifies papers for the reading list based on four inclusivity criteria described below. The total number of recommended papers is 20.

- Criteria 1: The list should comprise of a diverse set of papers that cover the different sub-topics related to the research topic

- Criteria 2: The list should comprise of popular papers which typically have higher citation counts
- Criteria 3: The list should comprise of one or more literature survey/review papers as these papers offer a synthesized overview of the research topic
- Criteria 4: The list should comprise of few recently published papers since the latest research works done on the research topic are also to be considered

Task 2 is the next task where the researcher intends to find similar papers based on a particular set of papers, after going through the reading list. The process called Citation Chaining is performed by researchers where citations and references of papers are used to find similar papers. In Rec4LRW, similar papers are identified using a hybrid process involving textual data, paper metadata such as author-specified keywords, shared co-references and shared co-citations of papers. For an input set of at least five seed papers, the Rec4LRW system recommends a maximum of 30 similar papers.

Task 3 is the last task in the Rec4LRW system. The objective of this task is to shortlist papers from the reading list (RL), based on the article-type preference of the user. While executing this task in the system, please imagine that you are at the stage of writing papers. The article-type options provided by the Rec4LRW system are generic research paper, case study, conference full paper and conference poster. Typically, this task is performed by the researcher while writing manuscripts for submission to journals and conferences. The Rec4LRW system identifies different clusters (communities of papers) by extracting the references and citations of papers in the reading list (RL). From these clusters, top papers are recommended to the user. The recommendations count varies with the article-type preference. For instance, if the

article-type selected by the user is conference poster, the recommendations count is the lowest while it is highest for the generic research paper article-type. The recommended papers need not be necessarily cited in the manuscript since selection of papers for citation is based on the user's personal context. However, the recommended papers are important papers as per the system and they are potential candidates for citation in the manuscript. This task is meant to be practically useful while managing a big set of papers during actual research studies.

Terminologies used in Rec4LRW

System Reading List (RL): Reading list is the list of all papers that are read during literature review. Researcher keeps populating this list as he/she finds new papers relevant to the particular research topic.

Initial Reading List (IRL): The 'initial reading list' (IRL) is a subset of the reading list (RL). The initial reading list is produced by the Rec4LRW system as a part of task 1.

Seed Basket (SB): Seed basket (SB) is a basket comprising of a particular set of papers. This set of papers is used to find similar papers as a part of task 2. In the Rec4LRW system, whenever you add a paper to the seed basket (SB), the paper also gets added to the reading list (RL).

Article-Type: The article-type is the type of the article that the researcher intends to write. Examples of article-types include research paper, conceptual paper, case study, conference full paper, conference short paper, poster, demo paper, technical report, white paper, master/doctoral dissertation etc.

Dataset used in Rec4LRW System

The dataset used in the Rec4LRW system is an extract of papers from the ACM Digital Library (ACM DL) (<http://dl.acm.org/>), published before 2011. The total number of papers in the extract is 115885, comprising of papers from both journals and proceedings. Not all papers indexed in the ACM DL are part of this extract as many papers had to be removed due to incorrect or missing data. Some papers are not part of the extract provided to us by ACM. For example, all the papers from the journal ACM Transactions on Information Systems are not part of the extract as the data was not provided to us. As per the data from this extract, papers published between 2008 and 2011 are considered as recent papers.

Experiment Steps

IMPORTANT NOTE: The Rec4LRW system is optimized for viewing in Google Chrome. Please avoid using other web browsers. Please try to complete the experiment in a single continuous session as the tasks are sequential in order

Stage 1: Performing and Evaluating Task 1

- 1) In the Rec4LRW homepage, click on task 1 hyperlink to start the experiment.
- 2) In the task 1 webpage, select the research topic of your choice from the drop-down box.
- 3) Click on the ‘Generate Recommendations’ button.
- 4) After the system retrieves the recommendations for the selected research topic, go through the papers.

- 5) Scroll down to the end of the page to find the hyperlink 'Click here to start evaluation of this task'.
- 6) Click the hyperlink to start the evaluation.
- 7) In the evaluation segment, provide the username which has been assigned to you.
- 8) After answering all the questions, click on the button 'Submit Survey'.
- 9) Click on the hyperlink provided in the 'thank you' page to start the task 2.

Stage 2: Performing and Evaluating Task 2

- 1) In order to perform task 2, you need to re-run task 1 for selecting at least five seed papers. This is done by clicking the button 'Task 1 papers' in the task 2 webpage.
- 2) After the system retrieves the recommendations, use the checkbox provided at the left of paper to select the papers. If a paper is selected, the system provides a notification that the paper has been added. The current number of papers in the seed basket is also provided as part of notification. You also have to option of removing a selected paper by unchecking the checkbox.
- 3) Perform the above step until you have selected five papers. You can also select more than five papers. The selected papers will be part of a seed basket (SB). If you wish to view the papers in the seed basket, click on the 'View Seed Basket' on the top right corner of the webpage.
- 4) After selecting the papers for the seed basket (SB), click on 'Generate Recommendations' button to retrieve recommendations for task 2.
- 5) After the system retrieves the recommendations for the selected research topic, go through the papers.

- 6) Scroll down to the end of the page to find the hyperlink 'Click here to start evaluation of this task'
- 7) Click the hyperlink to start the evaluation.
- 8) In the evaluation segment, provide the username which has been assigned to you.
- 9) After answering all the questions, click on the button 'Submit Survey'.
- 10) Click on the hyperlink provided in the 'thank you' page to start the task 3.

Stage 3: Performing and Evaluating Task 3

NOTE: Task 3 is a lengthy task with multiple steps. Please pay close attention to each step. This task requires at least 30 papers in your reading list (RL) for execution. The papers in the reading list (RL) comprise of papers that are explicitly added to it, and also the papers that were added to the seed basket (SB) in Task 2. While executing this task, please imagine that you are at the stage of writing papers.

- 1) In order to start task 3, you need to re-run task 2 for selecting papers for the final reading list (RL). This is done by clicking the button 'Task 2 papers' in the current webpage
- 2) After the system retrieves the recommendations, use the checkbox provided at the left of each paper to select the papers. At any point, you can click on 'View Reading List' link at the top right corner of the webpage to see the current list of papers in the reading list.
- 3) After selecting sufficient number of papers for the reading list (RL), select the article-type of your choice from the drop-down box.

- 4) Click on 'Generate Recommendations' button to retrieve recommendations for task 3. The system checks if the number of papers in the reading list is at least 30. If it is less than 30, the system prompts to either redo step 1 to add more papers to the reading list or run task-1 by clicking on the button 'Task 1 papers' to add more papers to the seed basket. Next run task 2 by redoing step 1 to add more papers to the reading list.
- 5) After the system retrieves the recommendations for the selected article-type, go through the shortlisted papers.
- 6) Scroll down to the end of the page to find the hyperlink 'Click here to start evaluation of this task'.
- 7) Click the hyperlink to start the evaluation.
- 8) In the evaluation segment, provide the username which has been assigned to you.
- 9) After answering all the questions, click on the button 'Submit Survey'.
- 10) Take a screenshot of the survey completion page using your mobile-phone/tablet or take a print-out of this page. This is required for collecting the cash incentive.

For any clarifications, please mail Aravind at the email address
aravind002@e.ntu.edu.sg

Part C: Task 1 Evaluation Questionnaire

Your username _____

Please answer the below questions pertaining to this task. You have the option of scrolling up this page to view the recommendations whilst answering the questions.

1. The recommendation list....

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Is relevant to the research topic	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consists of a good spread of papers for the research topic	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consists of papers from different sub-topics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consists of interdisciplinary papers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consists of papers that appear to be popular papers for the research topic	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consists of a decent quantity of recent papers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consists of a good mix of diverse, recent, popular and literature survey papers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The papers in the recommendation list appear familiar to you	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The papers in the	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

recommendation list are unknown to you					
The recommendation list consists of some unexpected papers that you were not expecting to see	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The papers in the recommendation list are useful for reading at the start of your literature review	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
This is a good recommendation list, at an overall level	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
There is a need to further expand this recommendation list	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. Please select your satisfaction level for this recommendation list

- ☐ Very Satisfied
- ☐ Satisfied
- ☐ Neutral
- ☐ Dissatisfied
- ☐ Very Dissatisfied

4. From the displayed information, what features did you like the most?

5. Please provide your personal feedback about the execution of this task

Part D: Task 2 Evaluation Questionnaire

Your username _____

1. The recommendation list...

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Consists of papers that are similar to the papers in the seed basket	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consists of papers that have shared co-references and co-citations with the papers in the seed basket	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. The recommendation list...

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Is relevant to the research topic	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consists of a good spread of papers for the research topic	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consists of papers from different sub-topics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consists of interdisciplinary papers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consists of papers that appear to be popular papers for the research topic	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Consists of a decent quantity of recent papers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
--	-------------------	----------	---------	-------	----------------

	Disagree				Agree
The papers in the recommendation list appear familiar to you	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The papers in the recommendation list are unknown to you	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The recommendation list consists of some unexpected papers that you were not expecting to see	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The papers in the recommendation list are useful for reading during your literature review	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
This is a good recommendation list, at an overall level	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
There is a need to further expand this recommendation list	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. Please select your satisfaction level for this recommendation list

- ☐ Very Satisfied
- ☐ Satisfied
- ☐ Neutral
- ☐ Dissatisfied
- ☐ Very Dissatisfied

5. The feature of adding papers to the seed basket to generate similar paper recommendations is a useful feature

- ☐ Strongly Agree

- ☐ Agree
- ☐ Disagree
- ☐ Strongly Disagree
- ☐ Not Sure

6. From the displayed information, what features did you like the most?

7. Please provide feedback on the execution of this task

Part E: Task 3 and System Evaluation Questionnaire

Your username _____

1.

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
The shortlisted papers are relevant to my article-type preference	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The shortlisted papers are useful for inclusion in my manuscript	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The shortlisted papers comprises of important papers from my reading list	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The shortlisted list comprises of papers which I would definitely cite in my manuscript	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
This is a good shortlisted papers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

list, at an overall level					
There is a need to further improve this shortlisted papers list	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2.

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
I would like to see the feature of shortlisting papers from reading list based on article-type preference, in academic search systems and databases	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I would like to see the feature of managing reading list and seed basket papers between the three tasks in academic search systems and databases	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. From the displayed information, what features did you like the most?

4. Please provide feedback on the execution of this task

Part F: Overall Evaluation of the Rec4LRW System

5.

	Strongly	Disagree	Not	Agree	Strongly
--	----------	----------	-----	-------	----------

	Disagree		Sure		Agree
The system is convenient	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I have to invest a lot of effort in the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
It takes many mouse-clicks to use the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Using the system takes little time	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
It takes too much time before the system provides adequate recommendations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

6.

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
I would recommend the system to others	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Using the system is a pleasant experience	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The system is useless	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The system makes me more aware of my choice options	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I make better choices with the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

I can find better papers by using the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

7.

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
Using the system would enable me to accomplish tasks more quickly	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Using the system would improve my work performance	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Using the system would improve my productivity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Using the system would enhance my effectiveness on the work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Using the system would make it easier to do my job	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I would find the system useful in my work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8. Please provide feedback on the system. You can mention features that you liked in the system and also other features that could be added

9. Please provide the time taken to complete this evaluation experiment (in minutes)

Minutes _____

APPENDIX I: CATEGORIES FROM THE QUALITATIVE ANALYSIS OF PARTICIPANT FEEDBACK FROM STUDY II

Part A: Task 1 Evaluation

Preferred Aspect Categories

Preferred Aspect Category (%)
Information Cue Labels (41%)
Rich Metadata (21%)
Quality (Diversity) (13%)
Quality (Overall) (9%)
Quality (Recency) (4%)
Quality (Interdisciplinary) (3%)
Quality (Survey) (3%)
Word Cloud (2%)
Speed (2%)
Good Mix (2%)
Link to ACM DL (2%)

Critical Aspect Categories

Critical Aspect Category (%)
Good (24%)
Broad Topics not suitable (20%)
Limited dataset issue (7%)
Quality can be improved (6%)
Different algorithm required (5%)
More recent papers required (4%)
Free-Text Search Required (4%)
Wanted more highly cited papers (4%)
Number of recommendations is inadequate (3%)
Advance search and filtering features required (3%)

Critical Aspect Category (%)
UI needs to be improved (2%)
Too diverse (2%)
Sorting options required (2%)
Wanted more recent papers (2%)
Ranking of papers not understandable (1%)
Links from word cloud required (1%)
Facets required for getting more papers (1%)
Browsing features required (1%)
Wrongly assigned cue labels (1%)
Need to validate with other datasets (1%)
Ranking, impact factor, popularity of papers required (1%)
Not sure on logic behind labels (1%)
Links from metadata fields required (1%)
Grouping of similar studies required (1%)
User control on number of recommendations (1%)
Interdisciplinary label required (1%)
Ranking based on citation count required (1%)
Evaluation window could have been in a new window (1%)
Grouping of papers required (1%)
More recent paper required (1%)

Part B: Task 2 Evaluation

Preferred Aspect Categories

Preferred Features Category (%)
Shared co-citations and co-references (28%)
Recommendation Quality (27%)
Information Cue Labels (16%)
Seed Basket (14%)
Rich Metadata (9%)
Ranking of articles (3%)
Speed (2%)

Preferred Features Category (%)
Good (1%)

Critical Aspect Categories

Critical Aspect Category (%)
Good (37%)
Quality can be improved (16%)
Limited dataset issue (12%)
Recommendation Algorithm could include more dimensions (7%)
Speed can be improved (7%)
Repeated recommendations from Task 1 (3%)
Sorting options required (3%)
Free-text searching required (3%)
User control on number of recommendations (3%)
Experiment procedure was long (3%)
Issue with understanding relations between results and SB papers (1%)
More recommendations could be provided (1%)
Too much information (1%)
Lesser papers in SB could be better (1%)
Not able to understand the ranking (1%)
Need to validate with other datasets (1%)

Part C: Task 3 Evaluation

Preferred Aspect Categories

Preferred Features Category (%)
Shortlisted papers (24%)
Information Cues (15%)
Cluster (11%)
Metadata (7%)
Seed Basket and Reading List (5%)
System Speed (2%)
UI (2%)

Critical Aspect Categories

Critical Aspect Category (%)
Good (28%)
Rote selection of papers for task execution (16%)
Limited dataset (5%)
Algorithm can be improved (5%)
Not sure of the usefulness (4%)
UI can be improved (3%)
Ranking can be improved (2%)
Speed can be improved (1%)
More user control needed (1%)
Free text searching (1%)
User instructions were initially not clear (1%)
More recommendations required (1%)
Not sure of the execution logic (1%)

Part D: System-level Evaluation

System Aspect Categories

System Aspect Category (%)
Good (16%)
Free-Text Search Required (13%)
Broad Topics not suitable (8%)
Limited Dataset hampered the quality (6%)
Seed basket is Good (5%)
Information Cues are good (4%)
Sorting options required (3%)
UI can be improved (3%)
Advanced search features required (3%)
Cluster option is Good (3%)
User control to set number of recommendation required (1%)
Download citations and data of paper (1%)
Need to validate the system with other datasets (1%)

System Aspect Category (%)
Task 1 is Good (1%)
Algorithms need to be improved (1%)
Experiment Procedure was Overwhelming (1%)
Task 3 is Good (1%)
Many user clicks in the experiment (1%)
More recent paper required (1%)
More similar papers needed in Task 2 (1%)
Social networking features can be included (1%)
Task 2 is Good (1%)
System needs to learn by itself (1%)
Explicit link between tasks need to be provided (1%)
Article-type of papers could be displayed (1%)
Facets required in UI (1%)
Speed needs to be improved (1%)
Fast (1%)
Import user data from other academic databases to start process automatically (1%)
User-level customization required (1%)
Can the algorithm be made opensource? (1%)
Differentiate results by finding out user's experience level (1%)
Diversity aspect is Good (1%)
Not able to find differentiation between different article-type recommendations (1%)
Grouping of similar papers required (1%)
Not able to see the benefit of the system from current systems (1%)
System can be used for narrowing down from broad to narrow topic (1%)
Not better than ACM and IEEE (1%)
Tag for interdisciplinary papers required (1%)
Not sure about background algorithms (1%)
Task 2 algorithm needs to be improved (1%)
Not sure whether the system can provide better results than human' (1%)
Task 2 speed needs to be improved (1%)
Number of seed basket and reading list could have been less (1%)

System Aspect Category (%)
Textual data in cluster could have been displayed as networks (1%)
Prefer Expert Advise (1%)
Limited papers in seed basket (1%)
Ranking of journals and conferences could be displayed (1%)
Ability to upload Bibtex file into seed basket (1%)
More survey papers required (1%)

APPENDIX J: LIST OF PUBLICATIONS

Periodicals

Raamkumar, A. S., Foo, S., & Pang, N. (2016). Survey on inadequate and omitted citations in manuscripts: a precursory study in identification of tasks for a literature review and manuscript writing assistive system. *Information Research*, 21(4). Retrieved from www.informationr.net/ir/21-4/paper733.html

Raamkumar, A. S., Foo, S., & Pang, N. (2017). Using author-specified keywords in building an initial reading list of research papers in scientific paper retrieval and recommender systems. *Information Processing & Management*, 53(3), 577-594. <https://doi.org/10.1016/j.ipm.2016.12.006>

Sesagiri Raamkumar, A., Foo, S., Pang, N. (2017). Evaluating a threefold intervention framework for assisting researchers in literature review and manuscript preparatory tasks. *Journal of Documentation*, 73(3), 555-580. <https://doi.org/10.1108/JD-06-2016-0072>

Sesagiri Raamkumar, A., Foo, S., Pang, N. (2017). User Evaluation of a Task for Shortlisting Papers from Researcher's Reading List for Citing in Manuscripts. *Aslib Journal of Information Management*, 69(6), 740-760. <https://doi.org/10.1108/AJIM-01-2017-0020>

Sesagiri Raamkumar, A., Foo, S., Pang, N. (2018). Can I have more of these please? Assisting researchers in finding similar research papers from a seed basket of papers. *The Electronic Library*. Manuscript has been accepted for publication

Proceedings

Sesagiri Raamkumar, A., Foo, S., & Pang, N. (2015). Rec4LRW-scientific paper recommender system for literature review and writing. *Frontiers in Artificial Intelligence and Applications* (Vol. 275). <https://doi.org/10.3233/978-1-61499-503-6-106>

Raamkumar, A. S., Foo, S., & Pang, N. (2015). Comparison of techniques for measuring research coverage of scientific papers: A case study. In *Digital Information Management (ICDIM), 2015 Tenth International Conference on* (pp. 132-137). IEEE. <https://doi.org/10.1109/ICDIM.2015.7381881>

Raamkumar, A. S., Foo, S., & Pang, N. (2015). More Than Just Black and White: A Case for Grey Literature References in Scientific Paper Information Retrieval Systems. In *International Conference on Asian Digital Libraries* (pp. 252-257). Springer, Cham. https://doi.org/10.1007/978-3-319-27974-9_26

Sesagiri Raamkumar, A., Foo, S., & Pang, N. (2016,). Making Literature Review and Manuscript Writing Tasks Easier for Novice Researchers through Rec4LRW System. In *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries* (pp. 229-230). ACM. <https://doi.org/10.1145/2910896.2925445>

Sesagiri Raamkumar, A., Foo, S., & Pang, N. (2016). What papers should I cite from my reading list? User evaluation of a manuscript preparatory assistive task. In *Proceedings of the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL2016)* (pp. 51–62). Retrieved from <http://ceur-ws.org/Vol-1610/paper7.pdf>

Technical Report

Raamkumar, A. S., Foo, S., & Pang, N. (2016). A Framework for Scientific Paper

Retrieval and Recommender Systems. *arXiv preprint arXiv:1609.01415*.

Retrieved from <https://arxiv.org/abs/1609.01415>