

User evaluation of a task for shortlisting papers from researcher' s reading list for citing in manuscripts

Sesagiri Raamkumar, Aravind; Foo, Schubert; Pang, Natalie

2017

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*, in press.

<https://hdl.handle.net/10356/85539>

<https://doi.org/10.1108/AJIM-01-2017-0020>

© 2017 Emerald. This is the author created version of a work that has been peer reviewed and accepted for publication by *Aslib Journal of Information Management*, Emerald. It incorporates referee' s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<http://dx.doi.org/10.1108/AJIM-01-2017-0020>].

Downloaded on 17 Apr 2021 17:53:21 SGT

User Evaluation of a Task for Shortlisting Papers from Researcher's Reading List for Citing in Manuscripts

Aravind Sesagiri Raamkumar¹, Schubert Foo, Natalie Pang

Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore

Abstract

Purpose – Although many interventional approaches have been proposed to address the apparent gap between novices and experts for literature review (LR) search tasks, there have been very few approaches proposed for manuscript preparation (MP) related tasks. This paper describes a task and an incumbent technique for shortlisting important and unique papers from the reading list of researchers, meant for citation in a manuscript.

Design/methodology/approach - A user evaluation study was conducted on the prototype system which was built for supporting the shortlisting papers (SP) task along with two other LR search tasks. A total of 119 researchers who had experience in authoring research papers, participated in this study. Online questionnaire was provided to the participants for evaluating the task. Both quantitative and qualitative analyses were performed on the collected evaluation data.

Findings – Results indicate that graduate research students prefer this task more than research and academic staff. The evaluation measures relevance, usefulness and certainty were identified as predictors for the output quality measure 'good list'. The shortlisting feature and information cues were the preferred aspects while limited dataset and rote steps in the study were ascertained as critical aspects from the qualitative feedback of the participants.

Originality/value – Findings point out that researchers are clearly interested in this novel task of shortlisting papers from the final reading list prepared during literature review. This has implications for digital library, academic databases and reference management software where this task can be included to benefit researchers at the manuscript preparatory stage of the research lifecycle.

Keywords - Manuscript preparation; Shortlisting citations; Citation recommendation; Digital libraries; Scientific paper recommender systems; Citation networks

Type – Research Paper

¹ Corresponding Author

Email addresses: aravind002@ntu.edu.sg (A. Sesagiri Raamkumar), sfoo@ntu.edu.sg (S. Foo), nlpang@ntu.edu.sg (N. Pang)

1 Introduction

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. The term “Scientific Publication Lifecycle (SPLC)” was used for referring to the lifecycle of a publication (journal article) in (Björk & Hedlund 2003). The term “Scholarly Communication Lifecycle” has been loosely used in certain studies (Harley et al. 2010; Wright et al. 2007) to highlight the different avenues of publication such as traditional journals, open-access journals and pre-print services. The research lifecycle schematic put forth in (Nicholas & Rowlands 2011) was 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) analyse research data, (vii) disseminate findings and (viii) manage the research process. Among these activities, the three major activities are literature review (LR), actual research work and dissemination of results through publication venues. These three activities comprise of multiple sub-activities that require specific expertise and experience (Levy & Ellis 2006). Information behaviour research has shown that researchers with low experience, face difficulties in completing research related activities (Du & Evans 2011; Karlsson et al. 2012). These researchers rely on assistance from supervisors, experts and librarians for learning the required skills to pursue such activities.

Apart from above mentioned process-based human interventions, technology-oriented interventions such as academic assistive systems have been built for alleviating the expertise gap between experts and novices in terms of research execution (Raamkumar et al. 2016). As a part of technology-oriented interventions, recommender systems (RS) techniques have been proposed to recommend scholarly information objects for a wide variety of tasks in the research lifecycle. Among the scholarly information objects, research papers are the most recommended objects in prior studies as papers are required for performing literature review and for ad-hoc information needs. Other information objects that have been recommended to researchers in prior RS studies include collaborators (Gunawardena & Weber 2009), co-authors (Sie et al. 2014) and publication venues (Beierle et al. 2016; Chen et al. 2015). In the research lifecycle, RS studies have majorly focused on the search tasks conducted during the LR stage. Such tasks include building an initial reading list at start of LR (Ekstrand et al. 2010; Jardine 2014), finding similar papers for seed paper(s) (Küçükünç et al. 2015; Huynh et al. 2012; Liang et al. 2011).

During the dissemination stage of the research lifecycle, researchers engage in manuscript preparation and writing for corresponding publication venues. For this phase, one unexplored area is helping researchers in identifying the important and unique papers that can be potentially cited in the manuscript. This identification is influenced by two factors. The first factor is the type of research for which citation of a particular paper makes sense due to the particular citation context [1]. The second factor is the type of article (for e.g., conference full paper, journal paper, and demo paper) that the author is intending to write. For the first factor, there have been previous studies (He et al. 2011; Huang et al. 2014; Shaoping 2010). The second factor represents a task that can be explored since

the article-type places a constraint on the citations that can be made in a manuscript, through dimensions such as recency, quantity, to name a few. This factor forms the focal point of the current study.

In this study, we seek to address the manuscript preparatory task of shortlisting papers from the final reading list (FRL) [2] of researchers based on article-type preference. We first describe the task and then propose a shortlisting technique for operationalizing the task. By the term ‘shortlisting’, we allude to the nature of the task in identifying important papers from the reading list. We implemented a recommendation technique for this shortlisting papers (SP) task as part of the functionality provided by an assistive system called Rec4LRW. This system was developed for providing recommendations for two other literature review tasks along with the SP task. The two LR tasks are (i) building an initial reading list of research papers and (ii) finding similar papers based on a set of papers. The system’s corpus of 103,739 articles and 2,320,345 references was built using an extract of ACM Digital Library (ACM DL).

A user evaluation study was conducted to evaluate the recommendations of the tasks and the overall Rec4LRW system, from researcher’s perspective. In this paper, we report the findings for the SP task from the evaluation study. The study was conducted with 119 participants comprising of graduate research students, academic staff and research staff. The participants, who had experience in publishing researcher papers, were selected through a screening survey. Participants had to select one of the provided 43 research topics and run the tasks offered by the system. During the study, the recommendations list, user-interface (UI) features and the overall system’s performance were evaluated by the participants through survey questionnaires. For the SP task, six evaluation measures were included in the evaluation questionnaire, to ascertain the agreement percentages from the participants. A multiple linear regression test was conducted for identifying significant predictors of the dependent measure ‘good list’ for the SP task. Subjective feedback of the participants was also collected through two questions in the questionnaire, for identifying thematic categories for the preferred features and the features that need to be improved.

The remainder of this paper is organized as follows. Section two surveys the related studies on application of RS in manuscript preparatory and writing tasks. In section three, the shortlisting technique of the task is explained. The Rec4LRW system is introduced along with dataset, technical details and UI features in section four. Details about the user study and data collection are outlined in section five. The evaluation results are presented and discussed in sections six and seven respectively. The concluding remarks and future plans for research are provided in the final section. The acronyms used in this paper are listed in Table I along with the corresponding descriptions.

Table I Acronyms used in this paper

| Acronym | Description |
|---------|----------------------|
| AP | Agreement Percentage |
| FRL | Final Reading List |
| IRL | Initial Reading List |
| LR | Literature Review |

| | |
|----|------------------------|
| MP | Manuscript Preparation |
| RL | Reading List |
| RS | Recommender System |
| SB | Seed Basket |
| SP | Shortlisting Papers |

2 Related Work

The studies surveyed in this section have been grouped under three stages of the manuscript preparatory and writing phase of the research lifecycle. The three stages and the constituent recommendation tasks are illustrated in Figure 1.

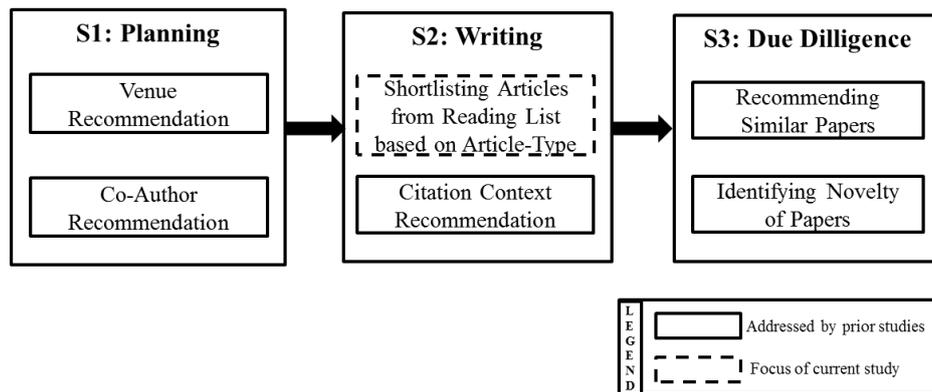


Figure 1 Stages and corresponding recommendations tasks in the manuscript preparatory and writing phase

Stage 1 (S1) refers to the stage where researchers are planning to write manuscript(s) for reporting their research findings. For this stage, two sets of RS studies have been conducted. The first set of studies is about recommending potential publication venues for the manuscript. 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. Collaborative filtering (CF) algorithms were used in a study where the focus was to help novice researchers (Yang & Davison 2012). The technique proposed in this study compared 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 was used to create a model based on the most frequent conferences attended by the co-authors. A similar approach was used in a recent study (Beierle et al. 2016) which made 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 et al. 2015) which uses the random walk algorithm for the recommendation process. The second set of studies in S1 concentrate on recommending co-authors for the manuscript. The co-authorship network data from historical publications was used a source for recommending co-authors with similarity computed using betweenness centrality in (Sie et al. 2012;

Sie et al. 2014). New co-authors were recommended using topical similarity between author-specified keywords of prior research papers and based on similar research interests.

Stage 2 (S2) refers to the stage where researchers are in the process of writing research papers. Citation context based RS studies fall under this stage. Generating recommendations for citation contexts is an approach meant to help the researcher in finding candidate citations for particular placeholders (locations) in the manuscript. These studies make use of content oriented recommender techniques as there is no scope for using collaborative filtering (CF) based techniques due to lack of user ratings. Translation models were used in (Huang et al. 2012; Lu et al. 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 technique in (He et al. 2011) is more author-friendly, as it does not expect the author to mark the citation contexts in the input paper unlike other studies where the contexts have to be set by the user. The proposed model in the study learns the placeholders in previous research articles where citations are widely made so that the citation recommendation can be made on occurrence of similar patterns. Multi-layer neural networks were used in (Huang et al. 2015) for learning the distributed semantic representations of the candidate words and the cited documents in the corpus. Due to enhanced learning, this technique is currently the best performing technique in citation context RS studies.

Stage 3 (S3) is the final stage for manuscript authors. This stage is the due diligence stage since authors attempt to enhance the quality of the manuscript. This stage has been addressed in two sets of studies. In the first set, the manuscript draft is taken as the input for generating recommendations. The Refseer system (Huang et al. 2014) provides both citation context recommendations and query text based recommendations. The topics from the user query (manuscript draft) are first extracted and citations are subsequently recommended using the topic-citation distribution from the system. An LDA model (He et al. 2010) is used for generating the topic distributions in this approach. In the second set of studies, novelty of a given manuscript is ascertained through the referenced papers. In McNee's study (Mcnee 2006), this scenario was addressed with two sub-tasks 'Explore novelty of a given paper' and 'Find more like this'. The usage of both user-based and item-based collaborative filtering algorithms was suggested for these sub-tasks. The citeomatic tool (A.I. 2017) is an operational tool developed to assist researchers in finding similar papers based on a set of references, particularly during the manuscript preparation stage. The citation prediction task in this tool is based on a deep learning model.

For the stage S2, even though article-type based recommendations have not been practically implemented, the prospective idea has been discussed in few studies. The article-type dimension has been highlighted as part of the user's 'purpose' in the multi-layer contextual model put forth in (Dehghani et al. 2011) and as one of the facets in document contextual information in (Champiri et al. 2015). 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 (RS) namely course recommendations (Winoto et al. 2012) and TV guide

recommendations (van Setten et al. 2006). Our work in the current study is the first to explore this task of providing article-type based recommendations with the aim of shortlisting important and unique papers from the cumulative reading list prepared by researchers during their literature review. Through this study, we hope to open new avenues of research which requires a different kind of mining of bibliographic data, for providing more relevant results to researchers.

3 Technique For Shortlisting Papers from Reading List

As mentioned earlier, the objective of this SP task is to help researchers in identifying important (based on citation counts) and unique papers from the final reading list (RL). These papers are to be considered as potential candidates for citation in the manuscript. For this task, the Girvan–Newman algorithm (Girvan & Newman 2002) was used for identifying the clusters in the citations network. The specific goal of clustering in this context is to identify the communities within the citation network. From the identified clusters, the top cited papers are shortlisted. The algorithm is implemented using the *EdgeBetweennessClusterer* method in JUNG library [3]. 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.

As a part of this task, we have tried to explore the notion of varying the count of shortlisted papers by article-type choice. For this purpose, four article-types were considered - conference full paper (cfp), conference poster (cp), generic research paper (gp) [4] 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 publications [5]. The number of papers to be shortlisted for these article-types was identified by using the historical data from ACM dataset. 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 process flow of the overall technique is illustrated in Figure 2. The detailed flowchart for this technique is provided in the Appendix section in Figure 8.

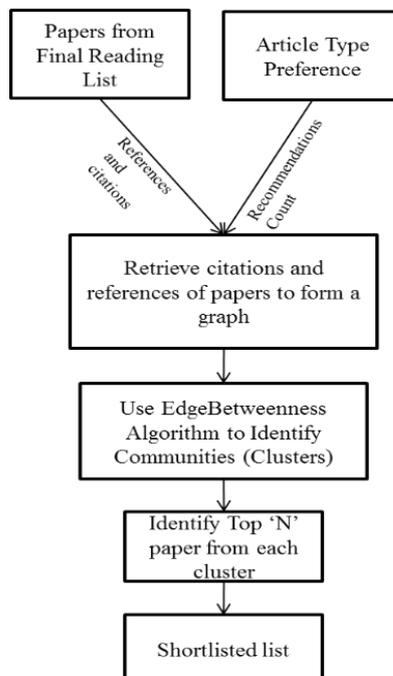


Figure 2 Process flow in the technique for shortlisting papers

4 Assistive System

The SP task was implemented as part of a system called Rec4LRW. This implementation was performed so that the task could be evaluated by researchers. In this section, a brief introduction of the Rec4LRW system is provided. The underlying dataset, task screens and the User interface (UI) features are also outlined.

4.1 Brief Overview

The Rec4LRW system was built as a tool aimed to help researchers (especially novices) in two main tasks of literature review and one manuscript preparatory task. The three tasks are (i) Task 1 - building an initial reading list of research papers, (ii) Task 2 - finding similar papers based on a set of papers, and (iii) Task 3 - shortlisting papers from the final reading list for inclusion in manuscript based on article-type choice. The usage context of the system is as follows. Typically, a researcher would run the first task for one or two times at the start of the literature review, followed by selection of few relevant seed papers which are subsequently used for Task 2. The second task consumes these seed papers as an input to find topically similar papers. This task is run multiple times until the researcher is satisfied with the whole list of papers. The third task (described in this paper), is meant to be run when the researcher is at the stage of writing manuscripts for publication. It is observed that the researcher would maintain numerous papers in his/her reading list while performing research (could be more than 100 papers for most research studies). The third task helps the researcher in identifying both important and unique papers from the full list of papers collated from Tasks 1 and 2. The shortlisted papers count is varied as per the article-type preference of the researcher.

The three tasks in the system are connected using two paper collection features called as seed basket (SB) and reading list (RL). 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. The papers to SB are added from Task 1 recommendations. Reading list (RL) is the list of all papers that are read during literature review. RL is different from the initial reading list (IRL) generated as part of Task 1 recommendations. RL is an operational feature implemented for the FRL concept. Researcher keeps populating RL as he/she finds new papers relevant to the particular research topic. Papers are added to RL under two scenarios. In the first scenario, whenever the user adds a paper to the seed basket (SB), the paper also gets added to the reading list (RL). In the second scenario, the user manually adds the paper to RL. This activity is performed after recommendations are generated for Task 2.

4.2 Dataset

A snapshot of the ACM Digital Library (ACM DL) was used as the dataset for the system. Papers from proceedings and journals for the period 1951 to 2011 form the dataset. The papers from the dataset were shortlisted based on full text and metadata availability in the dataset, to form the sample set/corpus for the system. The sample set contains a total of 103,739 articles and corresponding 2,320,345 references. The citation counts of the articles along with the other metrics were internally calculated.

4.3 User-Interface (UI) Features

In this sub-section, the UI features of the third task in the Rec4LRW system are presented. The input controls for this task is depicted in Figure 3. The article-type preference can be set from the drop-down list box. Along with the article-type, the main input to this task is the reading list (RL) papers that are accumulated from the first two tasks.

[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

STEP 1:

Click the below button to rerun task 2 for adding papers to your reading list

STEP 2:

Select the article-type of your manuscript

STEP 3:

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

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 3 Input controls for the third task in the Rec4LRW system

Shortlisted papers based on the article-type preference

- 1) [Measurement and analysis of online social networks](#) Popular
 Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, Bobby Bhattacharjee - Internet measurement, 2007
Abstract: Online social networking sites like Orkut, YouTube, and Flickr are among the most popular sites on the Internet. Understanding these graphs is important, both to improve current systems and to design new applications of online social networks.
Author Specified Keywords: analysis; measurement; social networks
 Citation Count: [115](#) References Count: [55](#) [View papers in the parent cluster](#)

Information cue labels
- 2) [SybilGuard: defending against sybil attacks via social networks](#) Popular
 Haifeng Yu, Michael Kaminsky, Phillip B. Gibbons, Abraham Flaxman - SIGCOMM Comput. Commun. Rev., 2006
Abstract: Peer-to-peer and other decentralized distributed systems are known to be particularly vulnerable to sybil attacks. In a sybil attack a malicious user obtains multiple fake identities and pretends to be multiple, distinct nodes in the system. By controlling a large fraction of the nodes in the system, the malicious user is able to "but vote" the honest users in collaborative tasks such as Byzantine failure defenses. This paper presents SybilGuard, a novel protocol for limiting the corruptive influences of sybil attacks. Our protocol is based on the "social network" among user identities, where an edge between two identities indicates a human-established trust relationship. Malicious users can create many identities but few trust relationships. Thus, there is a disproportionately-small "cut" in the graph between the sybil nodes and the honest nodes. SybilGuard exploits this property to bound the number of identities a malicious user can create. We show the effectiveness of SybilGuard both analytically and experimentally.
Author Specified Keywords: social networks; sybil attack; sybil identity; sybilGuard
 Citation Count: [40](#) References Count: [27](#) [View papers in the parent cluster](#)

Parent cluster of the shortlisted paper
- 3) [Structure and evolution of online social networks](#) Popular
 Ravi Kumar, Jasmine Novak, Andrew Tomkins - Knowledge discovery and data mining, 2006
Abstract: In this paper, we consider the evolution of structure within large online social networks. We present a series of measurements of two such networks, together comprising in excess of five million people and ten million friendship links, annotated with metadata capturing the time of every event in the life of the network. Our measurements expose a surprising segmentation of these networks into three regions: singletons who do not participate in the network; isolated communities which overwhelmingly display star structure; and a giant component anchored by a well-connected core region which persists even in the absence of stars. We present a simple model of network growth which captures these aspects of component structure. The model follows our experimental results, characterizing users as either passive members of the network; inviters who encourage offline friends and acquaintances to migrate online; and linkers who fully participate in the social evolution of the network.
Author Specified Keywords: graph evolution; graph mining; small-world phenomenon; social networks; stars
 Citation Count: [77](#) References Count: [30](#) [View papers in the parent cluster](#)
- 4) [Persona: an online social network with user-defined privacy](#) Recent
 Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, Daniel Starin - SIGCOMM Comput. Commun. Rev., 2009
Abstract: Online social networks (OSNs) are immensely popular, with some claiming over 200 million users. Users share private content, such as personal information or photographs, using OSN applications. Users must trust the OSN service to protect personal information even as the OSN provider benefits from examining and sharing that information. We present Persona, an OSN where users dictate who may access their information. Persona hides user data with attribute-based encryption (ABE), allowing users to apply fine-grained policies over who may view their data. Persona provides an effective means of creating applications in which users, not the OSN, define policy over access to private data. We demonstrate new cryptographic mechanisms that enhance the general applicability of ABE. We show how Persona provides the functionality of existing online social networks with additional privacy benefits. We describe an implementation of Persona that replicates Facebook applications and show that Persona provides acceptable performance when browsing privacy-enhanced web pages, even on mobile devices.
Author Specified Keywords: ABE; OSN; facebook; persona; privacy; social networks
 Citation Count: [12](#) References Count: [40](#) [View papers in the parent cluster](#)

Figure 4 Sample list of shortlisted papers for the task output

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 (RL). Apart from the regular fields such as author name(s), abstract, publication year and citation count, the system displays the fields:- author-specified keywords, references count and short summary of the paper (if the abstract of the paper is missing). Most importantly, we have included information cue labels beside the title for each paper. There are four labels (i) *popular*, (ii) *recent*, (iii) *high reach* and (iv) *survey/review*. A screenshot from the system for the cue labels (adjacent to article title) is provided in Figure 4.

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 was 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. The *high reach* label indicates that the paper has more number of references than most other articles for the research topic, thereby facilitating the scope for backward citation chaining. Specifically for Task 3, the system provides an option for the user to view the papers in the parent cluster of the shortlisted papers. This feature helps the user in serendipitously finding more papers for reading. The screenshot for this feature is provided in Figure 3.

5 User Evaluation Study

5.1 Objectives

In IR and RS studies, offline experiments are generally conducted for evaluating the proposed technique/algorithm with baseline approaches. Since the SP task addressed in the current study is a novel task, the best option was to perform a user evaluation study with researchers. The objective of the study was to investigate the usefulness and effectiveness of the task to researchers. In this evaluation exercise, models of information systems quality and success testing such as the TAM model (Venkatesh & Bala 2008) and IS success model (Delone & McLean 2003) were not employed as the current study's scope was restricted at the task level. The specific evaluation goals were (i) ascertain the agreement percentages of the evaluation measures, (ii) test the hypothesis that students are more benefitted from the recommendation tasks in comparison to staff, (iii) identify the significant predictors for the output quality measure Good List and (iv) identify the top preferred and critical aspects of the task from the subjective feedback of the participants. The nuances in handling the features of academic search systems along with task execution skills (Yoo & Mosa 2015; Niu & Hemminger 2012; Karlsson et al. 2012) are two of the factors which differentiate novice and expert researchers. These factors support our case for including hypothesis testing as one of the evaluation goals in this study. The claim is that students (novices) benefit most from the technological interventions which help them accomplish their literature search and manuscript preparatory tasks in an expedited manner.

5.2 Study Procedure

An online pre-screening survey was conducted to identify the potential participants. Participants needed to have experience in writing conference or journal paper(s) as a qualification for taking part in the study. All the participants were required to evaluate the three tasks and the overall system. In Task 1, the participants had to select a research topic from a list of 43 research topics. On selection of topic, the system provides the top 20 paper recommendations which were meant to be part of the initial LR reading list. In Task 2, they had to select a minimum of five papers from Task 1 into seed basket (SB). The system generated 30 topically similar papers for the selected SB papers. For the third task, the participants were requested to add at least 30 papers in the reading list (RL). The paper count 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 experiment 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 extract. The participant had to then select the article-type and run the task so that the system could retrieve the shortlisted papers. The detailed study guide provided to the participants can be accessed in this document [6].

5.3 Evaluation Measures

The evaluation measures in the study were selected based on the key aspects of the task. The selected measures *Relevance*, *Usefulness*, *Importance*, *Certainty*, *Good_List* and *Improvement_Needed* were meant to ascertain the quality of the recommendations. 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. *Relevance*, a frequently used measure, was included since it is important to identify whether the recommendations are relevant to the user input (for e.g. search keywords, images). In this study's context, the article-type preference is the main user input. Similarly, *Usefulness* is a measure meant to identify whether the output is useful for the user task (Ekstrand et al. 2010; Venkatesh & Bala 2008). In the current study, the user task is to identify papers for citing in manuscripts. *Importance* measure was for gauging the participants' perception on the relative importance of the recommended papers in milieu of the whole reading list of papers. In addition to the *Importance* measure, we wanted to determine participants' willingness to cite the papers using the *Certainty* measure. *Good_List* measure was adopted from an earlier study (Mcnee 2006) to ascertain the overall 'goodness of the list'. Even though, the aforementioned measures help in providing adequate indications of the output quality, we included the *Improvement_Needed* measure so that participants could explicitly indicate whether the recommendations need to be improved or not. Accordingly, the participants had to answer seven mandatory survey questions and two optional subjective feedback questions as a part of the evaluation. The seven survey questions and the corresponding measures are provided in Table II. A five-point Likert scale was provided for measuring participant agreement for each question. The values from the scale were strongly disagree (1), disagree (2), not sure (3), agree (4) and strongly agree (5). The evaluation questionnaire was provided to the user at the bottom of the Task 3 screen. The screenshot of the evaluation questionnaire is provided in Figure 5.

Table II Evaluation measures and corresponding questions

| Measure | Question |
|----------------------|---|
| Relevance | The shortlisted papers are relevant to my article-type preference |
| Usefulness | The shortlisted papers are useful for inclusion in my manuscript |
| Importance | The shortlisted papers comprises of important papers from my reading list |
| Certainty | The shortlisted list comprises of papers which I would definitely cite in my manuscript |
| Good_List | This is a good recommendation list, at an overall level |
| Improvement_Needed | There is a need to further improve this shortlisted papers list |
| Shortlisting_Feature | I would like to see the feature of shortlisting papers from reading list based on article-type preference, in academic search systems and databases |

Your username

1.

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

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

4. Please provide feedback on the execution of this task

Figure 5 Evaluation screen of the SP task

5.4 Analysis Procedures

The response values 'Agree' and 'Strongly Agree' were the two values considered for the calculation of agreement percentages (AP) for the evaluation measures. 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. For generating the predictive model, multiple linear regression (MLR) statistical test was employed. Statistical significance was set at $p < .05$. Statistical analyses were done using SPSS 21.0 and R. Participants' subjective feedback responses were coded by a single coder using an inductive approach (Thomas 2006), with the aim of identifying the central themes (concepts) in the text.

5.5 Participant Demographics

The study was conducted between November 2015 and January 2016. 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. 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 III for the different demographic variables. 62 participants were PhD/MSc students while 70 were research staff, academic staff and librarians. The average research experience for PhD students was 2 years while for staff, it was 5.6 years. Majority of the participants claimed they had

intermediate experience level (46.2%) and only a few participants claimed they were beginners (11.4%). 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.

Table III Demographic variables

| Demographic Variable | Number of Participants | |
|----------------------------|--------------------------|------------|
| <i>Position</i> | Student | 62 (47%) |
| | Staff | 70 (53%) |
| <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%) |

6 Results

6.1 Quantitative Data Analysis

6.1.1 Agreement Percentages (AP)

The agreement percentages (AP) for the seven measures by the participant groups are shown in Figure 6. In the current study, an agreement percentage above 75% is considered as an indication of higher agreement from the participants. As expected, the AP 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). It has been reported in earlier studies that graduate students generally look for assistance in most stages of research (Du & Evans 2011). Consequently, students would prefer technological interventions such as the current system due to the simplicity in interaction. Hence, the evaluation of students was evidently better than 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 APs. This observation validates the usefulness of the technique in identifying popular/seminal papers from the reading list. Due to favourable APs 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 behaviour. 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 studies.

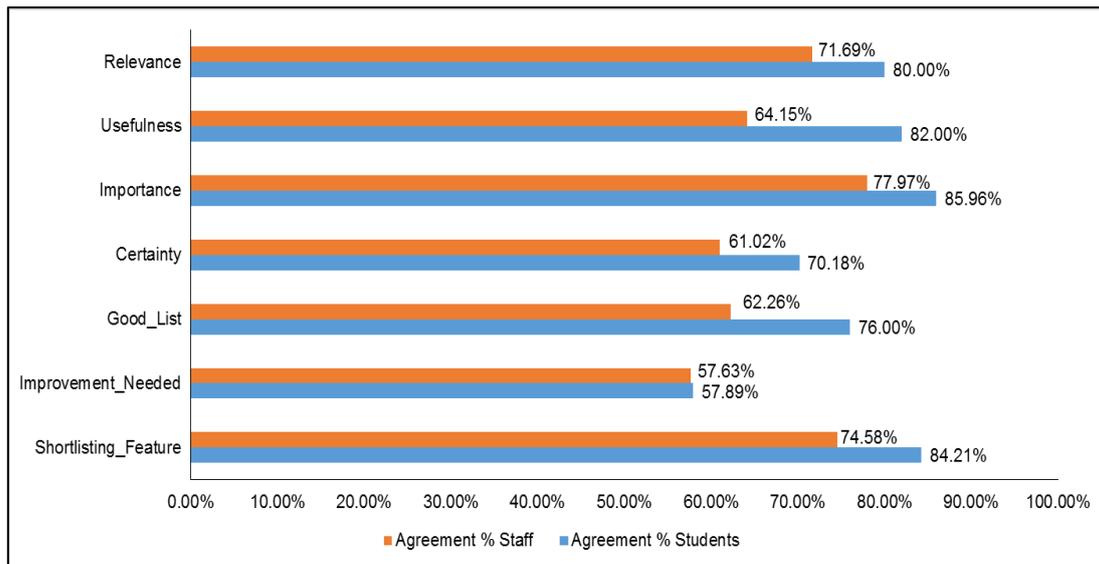


Figure 6 Agreement percentage results by participant group

6.1.2 Hypothesis Testing

Table IV lists the independent samples t-test results for the two groups. There was no statistically significant difference between the two groups for any of the seven measures at $p < 0.05$. Therefore, the hypothesis was not met for this task. 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 IV Independent samples t-test results

| Measure | t-value | Students | Staff |
|----------------------|---------|---------------|---------------|
| | | <i>M</i> (SD) | <i>M</i> (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) |

6.1.3 Regression Analysis

Results from multiple linear regression testing are displayed in Table V. 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. 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.

Table V Multiple linear regression results of Task 3

| | Estimate | SE | t value | p |
|----------------------|----------|------|---------|------|
| 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 *df*
Multiple R^2 : 0.722, Adjusted R^2 : 0.495
F-statistic: 19.820 on 6 and 109 *df*

6.2 Qualitative Data Analysis

In Table VI, the top five categories of the preferred aspects and critical aspects are listed. The percentages alongside the categories indicate the frequency of occurrence among the total set of comments for that task. In the following sub-sections, few examples of the participant feedback has been included for the categories.

Table VI Top five categories for preferred and critical aspects

| Rank | Preferred aspects categories | Critical aspects categories |
|------|---|---|
| 1 | Shortlisting Feature & Rec. Quality (24%) | Rote Selection of Papers (16%) |
| 2 | Information Cue Labels (15%) | Limited Dataset Issue (5%) |
| 3 | View Papers in Clusters (11%) | Quality can be Improved (5%) |
| 4 | Rich Metadata (7%) | Not Sure of the Usefulness of the Task (4%) |
| 5 | Ranking of Papers (3%) | UI can be Improved (3%) |

6.2.1 Preferred Aspects

24% of the participants felt that the feature of the shortlisting papers based on article-type preference was quite preferable and would help them in completing their tasks in a faster and efficient manner. They also felt that the quality of the shortlisting papers was satisfactory.

"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] [7]

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

15% of the participants felt that the information cue labels (*popular, recent, high reach and literature survey*) were helpful for them in relevance judgement of the shortlisted papers. This particular observation of the participants was echoed for the first two tasks of the Rec4LRW system, thereby validating the usefulness of information cue labels in academic search systems and digital libraries.

"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." - [R019]

"The tags associated with papers wherever they are present allows for easy classification." - [R58]

Around 11% of the participants felt the option of viewing papers in the parent cluster of the particular shortlisted papers was useful in two ways. Firstly, it helped in understanding the different clusters formed with the references and citations of the papers in the reading list. Secondly, the clusters served as an avenue for finding some useful and relevant papers in serendipitous manner as some papers could have been missed by the researcher during the literature review process. The other features that the participants commended were the metadata provided along with the shortlisted papers (citations count, article summary) and the paper management collection features across the three tasks.

"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]

6.2.2 Critical Aspects

Around 16% of the participants felt that the study procedure of adding 30 papers to the reading list (RL) as a precursor for running the task was uninteresting. The reasons cited were the irrelevance of some of the papers to the participants as these papers had to be added just for the sake of executing the task while some participants felt that the 30 papers count was too much while some could not comprehend why these many papers had to be added.

"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]

Around 5% of the participants felt that the study experience was hindered by the dataset not catering to recent papers (circa 2012-2015) and the dataset being restricted to computer science related topics. In our future studies, we plan to address this issue by including papers from other information sources such as CiteSeer and Microsoft Academic Research.

“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]

Another 5% of the participants felt that they shortlisting algorithm/technique could be improved to provide a better list of papers. A section of these participants needed more recent papers in the final list while others wanted papers specifically from high impact publications. Since this task and its incumbent technique is novel, there is scope for improving the algorithm.

“Showing the latest published top conference paper is important in computer science. Journal paper should have a lower priority compared to top conference paper, during the paper citation.” – [R50]

“It indeed lists popular relevant papers. It would be better if it shows the rank information of the conferences/journals.” – [R83]

7 Discussion

Evaluation results were promising in lieu of the task’s novelty in scientific paper RS 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 behaviour of researchers is very much based on their personal context. It is known that researchers mostly cite papers for supporting their personal claims and for evidence purposes (Erikson & Erlandson 2014). 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 many participants explicitly stated the usefulness of this task if incorporated in current academic search systems.

Among the UI features, certain novel informational display features in the Rec4LRW system were included for highlighting the aspects of the recommended papers. The effectiveness of these features was ascertained only through the subjective feedback of the participants. The information cue labels were appreciated 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 appreciated feature of ‘viewing papers in parent cluster’ was effective in its purpose. The feature showed why a particular paper has been shortlisted in particular clusters. 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 future 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. Overall, the

influence of the task recommendations was positively highlighted by the participants albeit with less intensity by expert researchers (staff).

For the benefit of future studies, there are a variety of contextual dimensions that can be considered for formulating recommendations in this task. The contextual dimensions identified for this task are illustrated in Figure 7. For the article-type dimension, the recommendations count has been varied in the current study for different article-types.

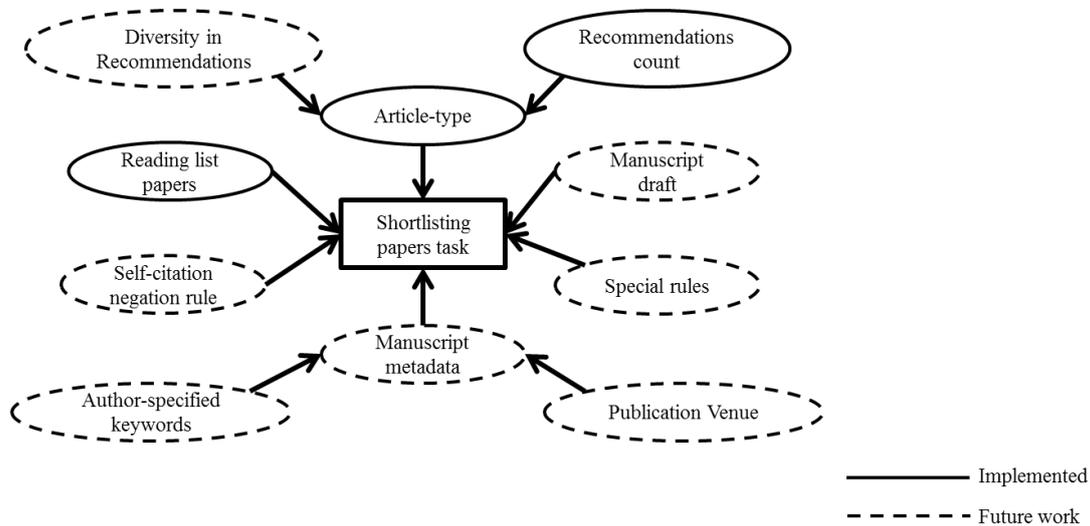


Figure 7. Contextual dimensions for the SP task

In addition, the type of recommendations could also be varied based on the article-type preference of the user. For instance, if the article-type chosen by the user is a conference full paper, the recommendations list could comprise of more recent papers since such papers are often cited in conference full papers. The intent is to diversify the recommendations based on the article-type. The manuscript metadata dimension is another important dimension for varying the recommendations. The author-specified keywords metadata field could be used to compute textual similarity with the reading list (RL) papers. Papers with high similarity values could be subsequently recommended. Similarly, the publication venue preference of the user could help in filtering certain papers. For instance, grey literature documents can be filtered out from the RL if the publication venue is a Tier-1 or top journal. The manuscript draft can be used as an input to the task, similar to the Refseer (Huang et al. 2014) and CiteSeerX systems (Wu et al. 2014). It is general practice that most researchers tend to avoid self-citation of their prior publications in manuscripts. This rule can be optionally enforced on this task. In addition, special rules for including or excluding certain papers based on researcher's personal preferences could be included in the task design. If all these contextual dimensions are considered for this task, the resultant recommendations would be of high quality.

8 Conclusion

For literature review and manuscript preparatory related tasks, the gap between novices and experts in terms of task knowledge and execution skills is well-known (Karlsson et al. 2012). A majority of the

previous studies have brought forth assistive systems that focus heavily on LR tasks, while only a few studies have concentrated on approaches for helping researchers during manuscript preparation. With the Rec4LRW system, we have attempted to address the aforementioned gap with a novel task for shortlisting papers from researcher's reading list, for inclusion in manuscript. The shortlisting task makes use of a popular community detection algorithm (Girvan & Newman 2002) for identifying communities of papers generated from the citations network of the papers from the reading list. Additionally, we have also tried to vary shortlisted papers count by taking the article-type choice into consideration.

In order to evaluate the system, a user evaluation study was conducted with 119 participants who had the experience of writing research papers. The participants were instructed to run each task. Participants were requested to answer survey questions and provide subjective feedback on the features of the tasks. As hypothesized before the start of the study, students evaluated the task favourably for all measures. There was high level of agreement among all participants on the availability of important papers among the shortlisted papers. This finding validates the aim of the task in identifying the papers that manuscript reviewers would expect to be cited. The predictors for the output quality measure 'agreeability on a good list' were identified. In the qualitative feedback provided by the participants, majority of the participants preferred the idea of shortlisting papers and also thought the output of the task was of good quality. Secondly, they liked the information cue labels provided along with certain papers, for indicating the special nature of the paper. As a part of critical feedback, participants felt that the study procedure was a bit longwinded as they had to select 30 papers without reading them, just for running the task.

As a part of future work, the scope for this task will be expanded to bring in more variations for the different article-type choices. For instance, research would be conducted:- (i) to ascertain the quantity of recent papers to be shortlisted for different article-type choices, (ii) include new papers in the output so that the user is alerted about some key paper(s) which could have been missed during literature review, (iii) provide more user control features in the system so that the user can select papers as mandatory to be shortlisted and (iv) integrate this task with the citation context recommendation task (He et al. 2011; Huang et al. 2014) so that the user can be fully aided during the whole process of manuscript writing.

Acknowledgments

This research was supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative and administered by the Interactive Digital Media Programme Office.

Notes

1. Citation contexts refer to the placeholders in a manuscript where the citation of papers is meaningful

2. Final reading list (FRL) refers to the collection of all research papers read by a researcher during the LR stage. On the other hand, initial reading list (IRL) refers to the set of research papers read by a researcher at the start of LR for acquiring a general overview of the research topic
3. Java Jung Library <http://jung.sourceforge.net/>
4. A paper is qualified as a generic research paper when it doesn't fall quality under the requirements of all the other article-types
5. Refer article classification in this webpage http://emeraldgrouppublishing.com/products/journals/author_guidelines.htm?id=oir
6. Rec4LRW user guide <http://goo.gl/dxUCuk>
7. RXX is a pseudonym pattern for the respondents in the evaluation study. R refers to respondent and XX refers to the number

References

- A.I., A., 2017. Citeomatic: Automated Literature Review. Available at: <http://allenai.org/semantic-scholar/citeomatic/> [Accessed April 7, 2017].
- 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*, pp.37–42. Available at: <http://dl.acm.org/citation.cfm?doid=2944789.2944871>.
- Björk, B.-C. & Hedlund, T., 2003. Scientific Publication Life-Cycle Model (SPLC). *ELPUB*. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.106.4255&rep=rep1&type=pdf> [Accessed February 25, 2014].
- 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), pp.1743–1758.
- Chen, Z. et al., 2015. Aver: Random Walk Based Academic Venue Recommendations. In *Proceedings of the 24th International Conference on World Wide Web*. pp. 579–584.
- Dehghani, Z. et al., 2011. A multi-layer contextual model for recommender systems in digital libraries. *Aslib Proceedings*, 63(6), pp.555–569.
- Delone, W.H. & McLean, E.R., 2003. The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, 19(4), pp.9–30. Available at: <http://www.tandfonline.com/doi/abs/10.1080/07421222.2003.11045748>.
- Du, J.T. & Evans, N., 2011. Academic Users' Information Searching on Research Topics: Characteristics of Research Tasks and Search Strategies. *The Journal of Academic Librarianship*, 37(4), pp.299–306. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0099133311000711>.
- Ekstrand, M.D. et al., 2010. Automatically Building Research Reading Lists. In *Proceedings of the fourth ACM conference on Recommender systems*. New York, New York, USA: ACM Press, pp. 159–166. Available at: <http://portal.acm.org/citation.cfm?doid=1864708.1864740>.

- Erikson, M.G. & Erlandson, P., 2014. A taxonomy of motives to cite. *Social Studies of Science*, 44(4), pp.625–637. Available at: <http://sss.sagepub.com/cgi/doi/10.1177/0306312714522871> [Accessed December 20, 2014].
- Girvan, M. & Newman, M.E.J., 2002. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12), pp.7821–7826.
- 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*. Available at: <http://www.aaai.org/ocs/index.php/FLAIRS/2009/paper/download/121</315%5Cnhttp://hdl.handle.net/1860/4145>.
- Harley, D. et al., 2010. Assessing the future landscape of scholarly communication: An exploration of faculty values and needs in seven disciplines. *Center for Studies in Higher Education*. Available at: <https://escholarship.org/uc/item/15x7385g.pdf> [Accessed September 9, 2016].
- He, Q. et al., 2011. Citation recommendation without author supervision. In *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11*. New York, New York, USA: ACM Press, pp. 755–764. Available at: <http://portal.acm.org/citation.cfm?doid=1935826.1935926>.
- He, Q. et al., 2010. Context-aware citation recommendation. *Proceedings of the 19th international conference on World wide web - WWW '10*, pp.421–430. Available at: <http://portal.acm.org/citation.cfm?doid=1772690.1772734>.
- Huang, W. et al., 2015. A Neural Probabilistic Model for Context Based Citation Recommendation. In *AAAI 2015: Proceedings of the Twenty-ninth AAAI Conference on Artificial Intelligence*. AAAI'15. AAAI Press, pp. 2404–2410. Available at: <http://dl.acm.org/citation.cfm?id=2886521.2886655>.
- Huang, W. et al., 2012. Recommending Citations : Translating Papers into References. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. pp. 1910–1914.
- Huang, W. et al., 2014. RefSeer : A Citation Recommendation System. In *Digital Libraries (JCDL), 2014 IEEE/ACM Joint Conference on*. pp. 371–374.
- Huynh, T. et al., 2012. Scientific publication recommendations based on collaborative citation networks. In *2012 International Conference on Collaboration Technologies and Systems (CTS)*. Ieee, pp. 316–321. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6261069>.
- Jardine, J.G., 2014. *Automatically generating reading lists*. Available at: <https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-848.pdf>.
- Karlsson, L. et al., 2012. From Novice to Expert: Information Seeking Processes of University Students and Researchers. *Procedia - Social and Behavioral Sciences*, 45, pp.577–587. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1877042812023312> [Accessed July 10,

2013].

- Küçükünç, O. et al., 2015. Diversifying citation recommendations. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(4), p.55:1-55:21. Available at: <http://dl.acm.org/citation.cfm?id=2668106> [Accessed March 9, 2016].
- Levy, Y. & Ellis, T.J., 2006. A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. *Informing Science: International Journal of an Emerging Transdiscipline*, 9(1), pp.181–212. Available at: <http://www.inform.nu/Articles/Vol9/V9p181-212Levy99.pdf>.
- Liang, Y., Li, Q. & Qian, T., 2011. Finding Relevant Papers Based on Citation Relations. In H. Wang et al., eds. *Web-Age Information Management: 12th International Conference, WAIM 2011, Wuhan, China, September 14-16, 2011. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 403–414. Available at: http://dx.doi.org/10.1007/978-3-642-23535-1_35.
- Lu, Y. et al., 2011. Recommending citations with translation model. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. New York, New York, USA: ACM Press, pp. 2017–2020. Available at: <http://dl.acm.org/citation.cfm?doid=2063576.2063879>.
- Mcnee, S.M., 2006. *Meeting User Information Needs in Recommender Systems*. Proquest. Available at: <http://search.proquest.com/docview/305306133>.
- Nicholas, D. & Rowlands, I., 2011. Social media use in the research workflow. *Information Services & Use*, 31(1–2), pp.61–83.
- 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), pp.336–353.
- 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). Available at: <http://www.informationr.net/ir/21-4/paper733.html> [Accessed December 20, 2016].
- van Setten, M. et al., 2006. Goal-based structuring in recommender systems. *Interacting with Computers*, 18(3), pp.432–456. Available at: <http://iwc.oxfordjournals.org/cgi/doi/10.1016/j.intcom.2005.11.005> [Accessed October 3, 2013].
- Shaoping, Z., 2010. *ActiveCite : An Interactive System for Automatic Citation Suggestion*. Available at: <http://scholarbank.nus.edu.sg/bitstream/handle/10635/22114/ZhouSP.pdf>.
- Sie, R.L.L. et al., 2014. COCOON CORE: CO-author REcommendations Based on Betweenness Centrality and Interest Similarity. In *Recommender Systems for Technology Enhanced Learning*. New York, NY: Springer New York, pp. 267–282. Available at: http://link.springer.com/10.1007/978-1-4939-0530-0_13 [Accessed October 18, 2016].
- Sie, R.L.L. et al., 2012. To whom and why should I connect? Co-author recommendation based on

powerful and similar peers. *International Journal of Technology Enhanced Learning*, 4(1–2), pp.121–137. Available at: <http://www.inderscience.com/link.php?id=48314> [Accessed October 17, 2016].

Tarjan, R., 1972. Depth-first search and linear graph algorithms. *SIAM journal on computing*, 1(2), pp.146–160.

Thomas, D.R., 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, 27(2), pp.237–246. Available at: https://flexiblelearning.auckland.ac.nz/poplhth701/8/files/general_inductive_approach.pdf [Accessed April 7, 2016].

Venkatesh, V. & Bala, H., 2008. Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, 39(2), pp.273–315. Available at: <http://doi.wiley.com/10.1111/j.1540-5915.2008.00192.x> [Accessed December 26, 2014].

Wilson, M.L., 2011. Search User Interface Design. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 3(3), pp.1–143. Available at: <http://www.morganclaypool.com/doi/abs/10.2200/S00371ED1V01Y201111ICR020> [Accessed March 16, 2016].

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), pp.56–75.

Wright, M. et al., 2007. Connecting digital libraries to eScience: the future of scientific scholarship. *International Journal on Digital Libraries*, 7(1), pp.1–4. Available at: <http://www.springerlink.com/index/832616v17076317m.pdf> [Accessed September 9, 2016].

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), pp.331–338.

Wu, J. et al., 2014. CiteSeerX: AI in a Digital Library Search Engine. Available at: <http://www.aaai.org/ocs/index.php/IAAI/IAAI14/paper/view/8607>.

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, pp.681–686.

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). Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4526974&tool=pmcentrez&rendertype=abstract>.

Zipf, G., 1949. Human Behavior and the Principle of Least Effort. In *Ed: Addison-Wesley*. Cambridge: MA: Addison-Wesley.

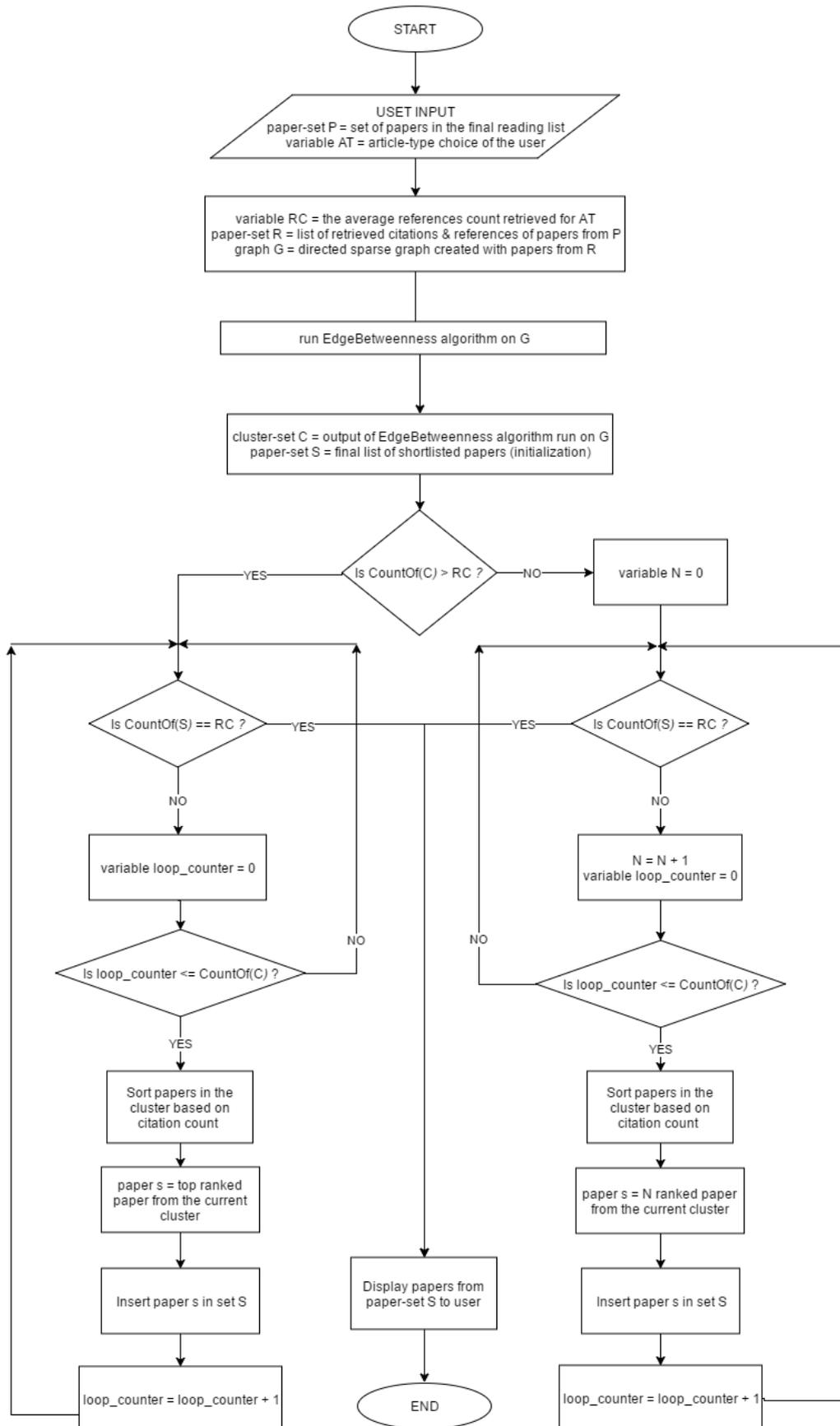


Figure 8. Flowchart of the shortlisting technique