# Making Literature Review and Manuscript Writing Tasks Easier for Novice Researchers through Rec4LRW System

Aravind Sesagiri Raamkumar, Schubert Foo, and Natalie Pang
Wee Kim Wee School of Communication and Information
Nanyang Technological University, Singapore
{aravind002, sfoo, nlspang}@ntu.edu.sg

## ABSTRACT

We demonstrate the recently built Rec4LRW system, meant for assisting researchers in three literature review and manuscript writing tasks. The system has been designed to be useful for all researchers, albeit the evaluation results show that it is more beneficial for research students and beginners. In this demonstration, we provide a walkthrough of the system by executing the tasks with sample research topics. The unique User-Interface (UI) and the task interconnectivity features are some of the highlighted aspects.

## Keywords

literature review; manuscript writing; reading list; shortlisting feature; scientific paper information retrieval; scientific paper recommender systems

## 1. INTRODUCTION

The gap between novices and experts in research related activities is an apparent phenomenon, due to lack of experience in terms of task knowledge and execution skills [1, 2]. For mitigating this situation, both process-based human and technological interventions have been proposed. Most of the technological interventions are piecemeal approaches with researchers having to depend on multiple and disparate avenues for assistance. With our literature review and manuscript writing assistive system, we aim to address aspects such as task interconnectivity, information cues and serendipitous paper discovery along with the good quality recommendations.

## 2. REC4LRW SYSTEM

### 2.1 Brief Overview

The Rec4LRW system [3] has been built to help researchers in three main tasks of literature review and manuscript writing. The tasks are (1) Building an initial reading list of research papers, (2) Finding similar papers based on a set of papers, and (3) Shortlisting papers from the final reading list for inclusion in manuscript based on article-type preference of the researcher. The recommendation techniques of the tasks are based on seven criteria. These criteria represent the characteristics of the bibliography and its relationship with the parent scientific paper. The high level characteristics of the bibliography are captured using four criteria: References Count, Citations Count, Grey Literature Percentage and Coverage. The next set of criteria is meant for capturing the relations between the scientific paper and each reference in the bibliography of the paper. They are Recency, Textual Similarity and Specificity. These criteria are used in the recommendation/retrieval techniques as per the individual task's requirement.

In task 1, coverage is used for ranking the final list of 30 papers from the output of a Content-based (CB) recommender that retrieves top 200 matching papers for the selected research topic. In task 2, the outputs of item-based collaborative filtering algorithm, title-based similarity matching and Textual Similarity & Specificity based document filtering techniques are merged. The seed basket is used as input for the aforementioned techniques. The final list of 30 papers is then ranked based on citation count. In task 3, a community detection algorithm is used to identify clusters of papers formed with the references and citations of the papers from the reading list. The number of papers to be shortlisted from these clusters is decided based on the article-type preference of the user. All the tasks in the system are interconnected using two collections called as *seed basket* and *reading list*. The seed basket serves as a collection of seed papers essential for running task 2 while reading list is a running collection of papers from both task 1 and 2. The reading list is one of the inputs for task 3. It is hypothesized that the system will be highly beneficial to research students and also for researchers who are venturing into new research topics.

### 2.2 Dataset and Technical Details

An extract of the ACM Digital Library (ACM DL) is used as the dataset. The dataset comprises of papers from proceedings and journals for the period 1951 to 2011. The sample set for the experiment was formed by filtering papers based on full text and metadata availability in the dataset. The final sample set contained a total of 103,739 articles. The back-end database is MySQL and JAVA is the main programming language used. The seven criteria values for all the papers in the sample set were measured as a pre-processing step. Apache Lucene and Apache Mahout libraries are used for the IR and RS mechanisms. The network analysis library JUNG is used exclusively in task 3 for creating the graphs and also for implementing community detection algorithms.

### 2.3 User-Interface of the System

The three tasks in the system are meant to be executed in a sequential manner. However, the tasks can be re-run as per user requirement if the intention is to mix different research topics in the same seed basket and reading list. This scenario potentially happens for multi-disciplinary research studies. Screenshot of one of these three tasks is provided in Figure 1. Certain novel UI features such as information cue labels (all tasks), shared co-relations with seed basket papers (task 2) and clustered papers option (task 3) have been included in the user interface to help researchers in relevance judgement and faster decision making.


*JCDL'16*, June 19–23, 2016, Newark, NJ, USA.
ACM 978-1-4503-4229-2/16/06.
DOI: http://dx.doi.org/10.1145/2910896.2925445

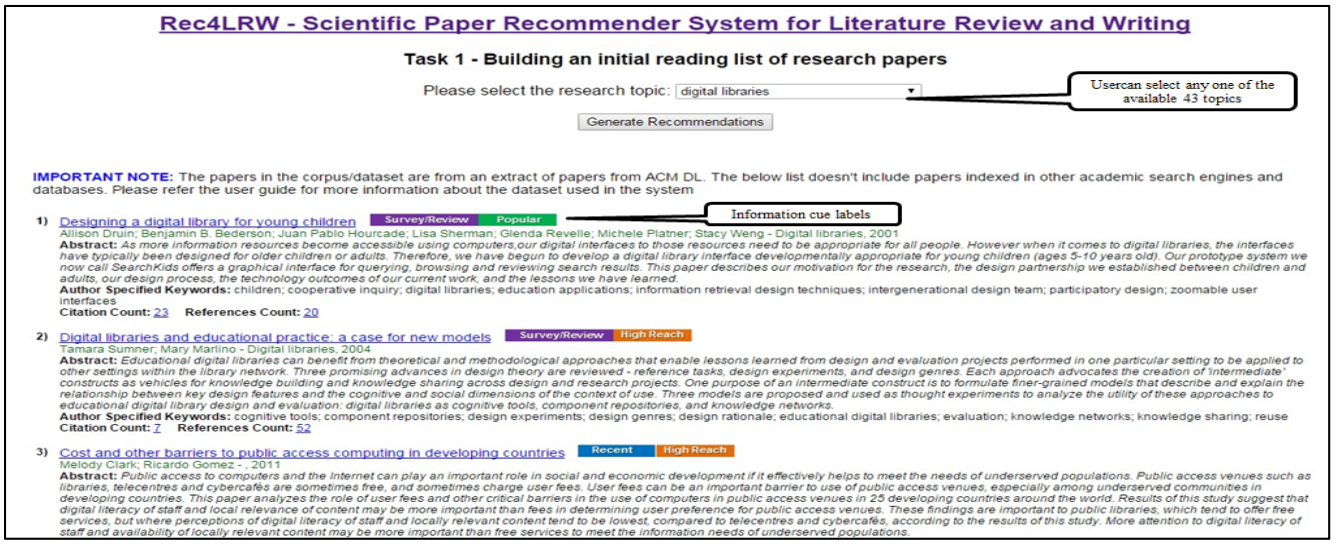


**Figure 1. Screen of Task 1 from Rec4LRW system**

## 3. USER EVALUATION STUDY

A user evaluation study was conducted to determine whether researchers using the tasks provided by Rec4LRW system can be efficient and effective in conducting the corresponding tasks in real-life settings. The participants had to select a research topic from a list of 43 provided topics. At the end of each task, participants had to answer mandatory survey questions and optional subjective feedback questions in evaluation questionnaires. System level evaluation questions were added to the questionnaire at end of task 3. Thee three system constructs chosen for the study are (i) Effort to use the System, (ii) Perceived System Effectiveness and (iii) Perceived Usefulness. As for the participants' demographics, 119 researchers participated in the whole study inclusive of the three tasks in the system. The reading list task (first task) was completed by 132 participants while 121 participants completed both the first and second task. 62 participants were PhD/MSc students while 70 participants were research staffs, academic staffs and librarians. The average research experience for PhD students was 2 years while for staffs, it was 5.6 years.

In Figure 2, the agreement percentages of the two study groups for the three system constructs are displayed. In the current study, an agreement percentage above 75% is considered as an indication of higher agreement from the participants. It is apparent from the percentages that students evaluated the system more favorably than staff. A significant difference of 14.18% was observed specifically for the Perceived Usefulness construct. This finding can be considered as an indication of the suitability of the system to students. The agreement is below 75% for the construct 'Effort to use the System' as participants found adding papers to the reading list as a rote activity performed solely for the study. The system's design rationale is to provide common recommendations for all users without taking user experience as a differentiating factor. However, students perceive the system to be more useful since it circumnavigates the apparent experience gap during execution of tasks.

## 4. FUTURE WORK

We plan to release a new version of the Rec4LRW system with more UI control features for enabling researchers to sieve through the results for better understanding the recommended papers. Additionally, we plan to set roles for the target users of the system so that personalization and customization features are made available. This feature is planned to be included in response to the qualitative feedback received from participants.

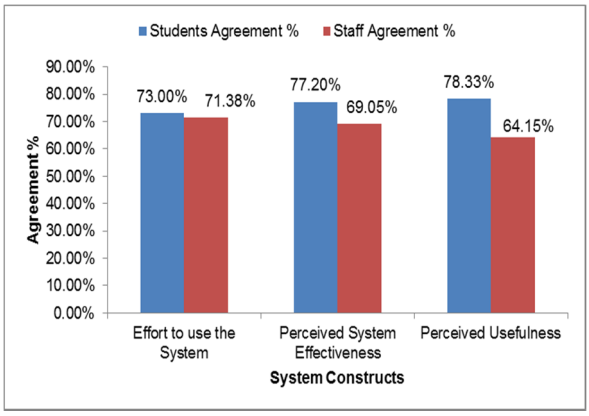


**Figure 2. Agreement percentages of the system evaluation constructs**

## 5. 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.

## 6. REFERENCES

[1] Du, J.T. and Evans, N. 2011. Academic Users' Information Searching on Research Topics: Characteristics of Research Tasks and Search Strategies. *The Journal of Academic Librarianship*. 37, 4 (Jul. 2011), 299–306.

[2] Karlsson, L. et al. 2012. From Novice to Expert: Information Seeking Processes of University Students and Researchers. *Procedia - Social and Behavioral Sciences*. 45, (Jan. 2012), 577–587.

[3] Sesagiri Raamkumar, A. et al. 2015. Rec4LRW – Scientific Paper Recommender System for Literature Review and Writing. *Proceedings of the 6th International Conference on Applications of Digital Information and Web Technologies* (2015), 106–120.