Supplementary File

# Introduction

This file contains additional information designed to supplement the paper titled “Coping with Large-scale, Ill-defined Literature: Developing a Tool to Enhance Searches”, submitted to the Journal of the Association for Information Science and Technology for consideration.

It consists of the following subsections:

Part A: An illustrative Use Case

Part B: Additional details in support of the Post-launch evaluation

(Bi) Workshop design,

(Bii) Survey design, and

(Biii) Participants.

# Part A: An Illustrative Use Case

This section narrates the journey of a research team’s experience of using the PermuSearch tool in their literature search efforts. The team consisted of 6 researchers (with a mix of experienced and novice researchers) dispersed across Asia and Australia with diverse backgrounds (Management Information Systems, Operations Research, and Marketing - none from an Information Science background). The primary aim was to Understand the status of constraint-based innovation literature. The research team aspired to collate all relevant literature in the domain to unpack trends (in the form of a meta-analysis) and to theorise the different innovation strategies that have been used for diverse constraints across diverse contexts.

As part of their literature searching efforts, the team engaged in a variety of seeking patterns:

*Surveying*: They wanted to understand the volume and spread (i.e., what journals and (sub-)disciplines the related papers came from) to design the study scope accordingly. This involved an iterative journey as the team recognised early on that there were much more related concepts than originally anticipated. Chaining: The research team intended to use inward and outward citations to unpack this further.

*Browsing*: They would browse through search results (reading their abstracts and titles) iteratively, unpacking new terms and re-constructing the search strategy each time.

*Distinguishing*: The team also paid attention to the outlets (journals and/or conferences) and their rankings, being peer reviewed or not, types of papers identified (e.g., empirical, conceptual etc.) when browsing the search results to form early impressions of the profile of the papers extracted . Once again, these insights influenced follow-up search iterations.

*Extracting and filtering*: The papers deemed relevant in this way were downloaded, and shared via a OneDrive Folder across the team, for their relevance and quality to be assessed, to determine inclusion and exclusion in the analysis. This was done by one team member - separate from the ones who executed the search and was set up as a quality assurance process.

*Monitoring*: Given this work was done to form the basis of a long-term research collaboration, the team was keen to set up alerts when new papers are available – but could not set this up systematically for the diverse sub-topics and search terms they had used.

The team accessed the PermuSearch tool at http://www.permusearch.com. Once in the landing page (see Figure A.1), one needs to download the template to specify the search parameters and also need to have an API key set up to start using the tool – both simple steps.

A screenshot of a computer

Description automatically generated

Figure A.1: PermuSearch landing page

The ‘Search’ tab in the template is where you specify the search parameters (see Figure A.2). You can set up to two dimensions (denoted as X- and Y dimensions in the sheet), where you can set search terms related to the *topic*, outlet (*venue*) and the *year* of publication, and also add up to three constraints – these specify fixed search parameters for each permutation of the search, such as based on a fixed topic, set of venues, or period.

A screenshot of a computer

Description automatically generated

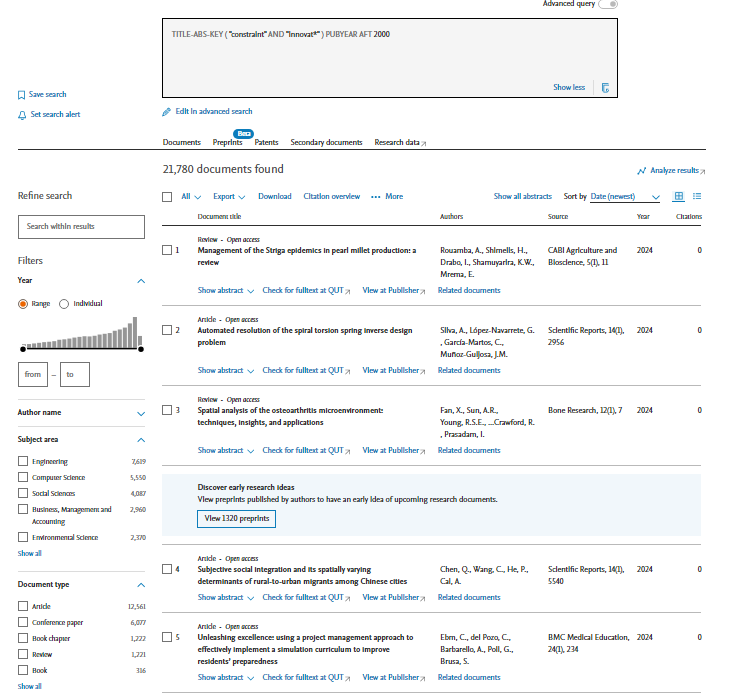
Figure A.2: Set Template

In this example, the research team aimed to identify all relevant papers after the year 2000 that were about constraint-based innovation. Forming the search terms was not straightforward, as it was anticipated that there would be many papers across diverse disciplines with diverse other terms used – and the team was not confident in identifying things upfront. Therefore, to obtain an initial feel for the current discourse of the topic, a simple search was executed referring to “Constraint-based” AND “innovat\*” as topical search terms specified for the X dimension and “PUBYEAR AFT 2000” as the only constraint. See Figure A.3.

A screenshot of a computer

Description automatically generated

1. Setting the search term



1. Reviewing the results

Figure A.3: Initial search done

This resulted in 21,780 publications across diverse disciplines and outlets, linked in the PermuSearch output to the listing of these publications in Scopus. Browsing through the results, it pointed to several ways the search scope can be improved.

*First*, the term “innovat\*” led to increased ‘false positives’ in the search results, for many papers would discuss an “innovative” approach (describing the overall research outcomes) that did not relate to innovation literature. This guided the team to change the search term.

*Second,* new terms that were proxies to constraint-based innovation were identified which was also supported by the backward and forward searching of papers deemed relevant. For example; ‘Frugal’, ‘Resource constrained’, ‘Resource scarcity’, ‘Bottom of pyramid’ and ‘Bricolage’ were terms close to ‘constraint-based’ - sometimes used interchangeably as well, in the literature. And ‘New product development’ was a seeming proxy to “Innovation”.

*Thirdly,* papers pertaining to some fields (e.g. medicine, physics, maths) were mostly irrelevant to the study context. As even through the search terms were used, they were used in very different contexts and meaning. (Note that PermuSearch can not directly help with scoping the domains- and this was done directly within Scopus].

These were taken as input to form the next search round. Figure A.4 shows the input file used here. Figure A.5 shows the results, illustrating the most prevalent terms from the list in yellow highlights.

A screenshot of a computer

Description automatically generated

Figure A.4 Input for Second round

A white background with black and yellow text

Description automatically generated

Figure A.5 Results from Second round

In order to confirm that the prevalent terms identified were covering similar (and relevant) topics, the team also ran a ‘cross topic’ search (entering the terms in Dimensions X and Y - see Figure A.6). This allowed the team to see how often one set of terms were used in papers that also had another set of terms (Figure A.7)- hence increasing the researcher’s confidence in selecting the main search terms to include in their search strategy.

A screenshot of a computer

Description automatically generated

Figure A.6 Setting up the cross-term search

A screen shot of a computer

Description automatically generated

Figure A.7 Results from the cross-term search

As Figure A.7 depicts, some terms such as “Bottom of the Pyramid” and “New Product development” (see greyed areas) were not very prevalent. However, the cases where these terms were used with other similar terms (e.g. the 4 cases of ["Frugal" AND "New product development" AND ("constraint" AND "innovation\*"), resulted in highly relevant papers - however few the numbers were. The more prevalent combinations (see pink and orange highlights) also resulted in relevant papers. Thus, the cross-topic search serving as an expeditor to the paper relevance checking process. It also helped to identify search terms that might be nested (see green area). In this case, the search term ("resource constrain\*" AND "innovation\*") produced a subset of the outcomes from the string ("constraint" AND "innovation\*")- thus pointing to unnecessary duplications that can be removed from the search terms.

***Summary Reflections***

As the search commenced, they were not at all confident on what terms to use (i.e., if they had captured all related areas that may be within scope). The iterative progression helped here and the ability to search with multiple terms in effortlessly while interfacing with Scopus – with the results (and thus, enjoy all the features Scopus already offered) made the searching a much more efficient process with a clean and clear trial of evidence. The PermuSearch outputs were easily sharable across the team members and helped much with obtaining multiple-researcher input within the search process.

# Part B: Additional details in support of the Ex-post evaluation

The ex-post evaluation of PermuSearch consisted of surveying a group of academics during a workshop in which they were introduced to the tool and had some time to use it for themselves.

## **B.1 Workshop design**

The workshop took place in-person, in a seminar room during an Information Systems conference in 2023. In it, participants (n=22) were first asked through a Qualtrics survey, to report their difficulties in searching the literature, and how long they had been in academia.

We then introduced PermuSearch, demonstrated it, and made it available to all participants. A Scopus API key was provided for use during the workshop, such that participants did not need to request one themselves. Participants were free to ask questions during this session.

After the coffee break, participants returned to the venue (n=13) and took part in another Qualtrics survey, which targeted the evaluation of PermuSearch itself.

## **B.2 Survey design**

The survey received Ethical Approval from the University of Otago. It consisted of two Qualtrics questionnaires. The survey was anonymous, with informed consent, and taken when instructed to do so during the workshop. It was accessed via a short URL and a QR code displayed on the screen. Figure B.1 shows the questions of the first questionnaire, and Figure B.2 shows those of the second questionnaire. The usefulness, ease of use, and satisfaction questions were sourced from the USE Questionnaire based on Lund (2001).

A screenshot of a survey

Description automatically generated  
*Figure B.1 Questions from Questionnaire 1*

A survey form with many circles

Description automatically generated A screenshot of a survey

Description automatically generated*Figure B.2 Questions from Questionnaire 2*

## **B.3 Participants**

Respondents to our first questionnaire were generally relatively new, but not inexperienced in academia, with over half of our participants having between 1 and 4 years of experience. Figure B.3 provides a breakdown.

Figure B.3 Years of academic experience of participants

Before introducing PermuSearch, we asked participants how often they used selected literature search tools, and their perceived difficulties in conducting literature searches, with results shown in Figure B.4. While Google Scholar was used more frequently, Scopus was used by all participants, indicating that any drilling down into the searches would be in a (at least somewhat) familiar environment.

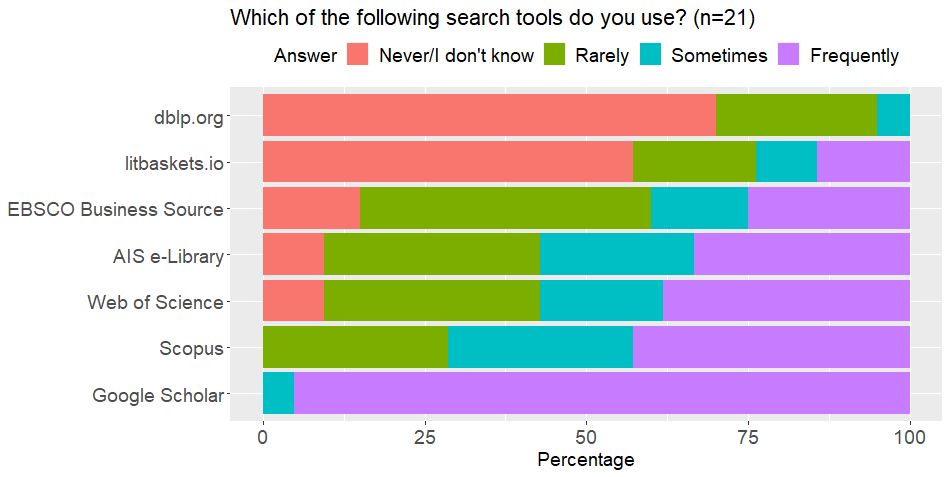


Figure B.4: Frequency of participants' use of selected literature search tools

# References

Lund, A. M. (2001). Measuring usability with the use questionnaire. *STC Usability SIG Newsletter*, *8*(2), 3-6. <https://garyperlman.com/quest/quest.cgi?form=USE>