COLLEGE OF INFORMATION SCIENCES AND TECHNOLOGY

THE PENNSYLVANIA STATE UNIVERSITY

Quantitative Analysis of Bibliography Management Applications

What Makes A Good Reference Manager?

Tongan Cai
cta@psu.edu

Chacha Chen
cjc6647@psu.edu

16TH DECEMBER, 2019

COLLEGE OF IST, UNIVERSITY PARK, PA 16802

CLASS PROJECT FOR IST 521

Quantitative Analysis of Bibliography Management Applications: What Makes A Good Reference Manager?

Tongan Cai cta@psu.edu

Chacha Chen cjc6647@psu.edu

16TH DECEMBER, 2019

Abstract

Reference managers, or bibliography management applications, are widely used in academia. However, it is always controversial to determine which bibliography management application is better in an objective way. While related work provides ideas of qualitative analysis, this work steps forward and adopts a quantitative approach to evaluate the reference managers by the amount of physical and mental effort users make in using bibliography management applications.

Specifically, besides qualitative comparison, the proposed quantitative analysis adopts a keystroke and mouse move logger—RUI to record and analyze the user's mouse and keyboard activities. Subjects are first asked with a pre-study survey about their favorite reference manager and their thoughts about the usability of reference management applications and then perform a specific task to search, collect and construct a bibliography list. Four common bibliography management applications—Mendeley, Zotero, EndNote, and RefWorks across both Windows and OS X operating systems. The process is also simulated with Cogulator[1] to model the mental effort that the user is making. A retrospective survey is asked if the subject will prefer another reference manager and if they are now aware of the usability issues.

Our results indicate that different reference managers do require different levels of effort from the users. With less effort made, the user is more likely to mark the bibliography management application as preferable. The users' discrepancy of preference comes from their unfamiliarity of applications they did not get a chance to try. We further conclude that though the applications share similar functionalities and features, the difference in their presentations matters. Factors like pricing, cloud sync and accuracy of bibliography generation also influence the preference of users. The proposed method is also expected to be applied in comparatively analyzing other applications, and transferred to other user-oriented analyses.

[1] http://cogulator.io/ the open-source application for human performance calculation

Table of Contents

1 Introduction

Reference managers, or reference management applications, bibliography management applications, are computer applications or systems that help users to collect, manage and format references and citations for academic purposes. It is both vital and tedious to manually collect and keep track of all citations, but with the help of reference managers, some scholars will feel comfortable and relieved. As Lorenzetti and Ghali (2013) mentioned, reference management applications are optimized and well used for writing literature reviews. The reference management applications also helps to make PDF files to be instantly viewable and searchable (Mead & Berryman, 2010).

Yet it is not always agreed which reference manager is the best, and different people have different preferences of reference managers. The website G2 shows that the most liked reference manager is Mendeley[2] with 170 reviews and 4.3/5 rating. On another website, Scribendi, RefWorks is the top pick from the editor[3]. On the other hand, the history of publicly available, web-based reference management applications started decades ago, but most of the reference management applications are still in an old-fashioned look (like Mendeley, Zotero). They are much goal-oriented, without taking consideration of usability (Mead & Berryman, 2010).

The comparison of reference managers can take reference from the assessment of usability. Various software quality models have been proposed to improve the usability of software (Al-Badareen, Selamat, Jabar, Din, & Turaev, 2011). There are more concrete methods in evaluating the usability of software. Park and Hwan Lim (1999) proposes that software ergonomics can have a quantitative basis with regard to multiple criteria of usability concern in comparing interfaces. A first prescreening phase takes expert judgments and the second evaluation phase involves user-based assessment. Bellamy, John, and Kogan (May 21, 2011) introduces a CogTool that analyzes tasks with an interactive system and predicts the time an expert will take to perform those tasks. An application of such a two-stage idea is presented in De Cristofaro, Du, Freudiger, and Norcie (2013), where a pre-study survey was conducted before a quantitative study on the usability of three popular two-factor authentication solutions.

Some other previous works conducted surveys to compare between different reference managers. Lorenzetti and Ghali (2013) conducts a survey on reference manager users about their usability. It emphasizes the ease-of-use issues concerning the integration of the software with other programs and the sharing of reference databases among researchers. Other works have metrics like error rate (Gilmour & Cobus Kuo, 2011) and import time (Santharooban & Lavakumaran, 2018). Basak (2014) quantified the field import by Google Scholar on RefWorks, Mendeley, and EndNote. However, few work considered the actual amount of effort for users to complete a routine task on reference managers, making their analysis more subjective.

For this study, we address this problem by conducting a quantitative, comparative study

[2]https://www.g2.com/categories/reference-management

[3]https://www.scribendi.com/advice/reference_management_software_solutions.en.html

among four commonly used reference managers–Mendeley, Zotero, EndNote, and RefWorks, aiming to distinguish different level of efforts for users to operate them. Specifically, we will reference and improve the ideas in previous works and conduct a quantitative analysis of user efforts by 1) survey on targeted subjects about the usage and thoughts on reference management applications, 2) compare the functionalities and features of each reference management application, 3) record and analysis the mouse movements and keystrokes in operating the application, and 4) build separate Key-stroke Level Models (KLM) with Cogulator for each of the applications.

The remaining text is organized as follows: Section 2 introduces the methods and materials of the study; Section 3 presents the experiment process and corresponding results, and Section 4 draws our conclusions and discuss the potentials of future works.

2 Methods

2.1 Subjects

We have two groups of subjects. The first group is for the pre-study survey, of which the participants are mostly graduate students or senior undergraduate students in the STEM field. A total of 24 subjects responded to the survey. The second group is for the quantitative study, where participants are experts in operating the designated reference management applications on both Windows and OS X operating systems. 5 subjects volunteered for the study. The selection of the first group is based on an online distribution of a survey link among multiple online chat groups. The subjects in the second group are selected from the first groups who agreed to participate in the optional, voluntary quantitative study. Demographics and personal information are not collected for the first group, and the second group consists of 3 male subjects and 2 female subjects. All of them are first-year graduate students.

2.2 Materials

The materials we have are:

- Desktop running Windows 10 (1920x1080 resolution)
- Desktop running OS X 10.14 (1920x1080 resolution)
- Mendeley Desktop v1.19.5 on Windows and OS X, with Chrome plug-in
- Zotero 5.0 on Windows and OS X, with Chrome plug-in
- EndNote X9 on Windows and OS X
- RefWorks by ProQuest
- RUI 2.3 on Windows and RUI 2.1 on OS X
- Cogulator v4.1.1 on OS X
- A short list of titles for academic papers

Titles listed include:

- *Generative Adversarial Imitation Learning*
- *Biomedical Image Sequence Analysis with Application to Automatic Quantitative Assessment of Facial Paralysis*
- *2D Landmark-Based Facial Asymmetry Assessment in the Clinical Case of Facial Paralysis*

2.3 Reference Management Applications

We first review the reference management application we're studying for this project: Mendeley, Zotero, EndNote and RefWorks.

2.3.1 Mendeley

Mendeley is a free reference manager by Elsevier. The Mendeley is supported on various platforms, including mobile phones and tablets. A screenshot of the Mendeley's interface is shown in Figure 1. The UI is quite clean and intuitive while supporting full functionalities.

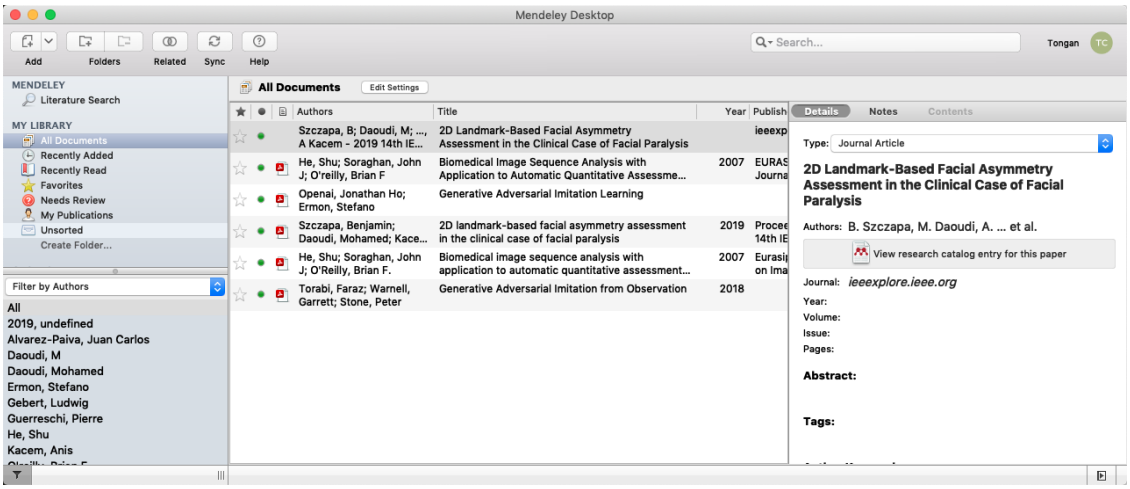


Figure 1: A screenshot of Mendeley

Mendeley allows users to have their personal account as well as a browser plug-in that can capture publication information on a webpage. If full text is available, Mendeley will also store the PDF files of the paper. The Mendeley owns a hugh group of users. In 2018, it was estimated that there are already 5 million users for Mendeley[4].

Some issues regarding Mendeley are the relatively low query speed of browser plug-in and weak auto-completion of information. For the relatively new conference paper *Generative Adversarial Imitation Learning*, the Mendeley fails to find its information from its database and provides with an unpublished version.

[4]https://www.elsevier.com/connect/ten-years-of-mendeley-and-whats-next

2.3.2 Zotero

Zotero is also a free, open-source reference management application by the Corporation for Digital Scholarship. Zotero has web browser integration, online syncing, generation of in-text citations, footnotes, and bibliographies and integration with the Microsoft Word and Google Docs. A screenshot of the Zotero's interface is shown in Figure 2. It supports similar functionalities as Mendeley.

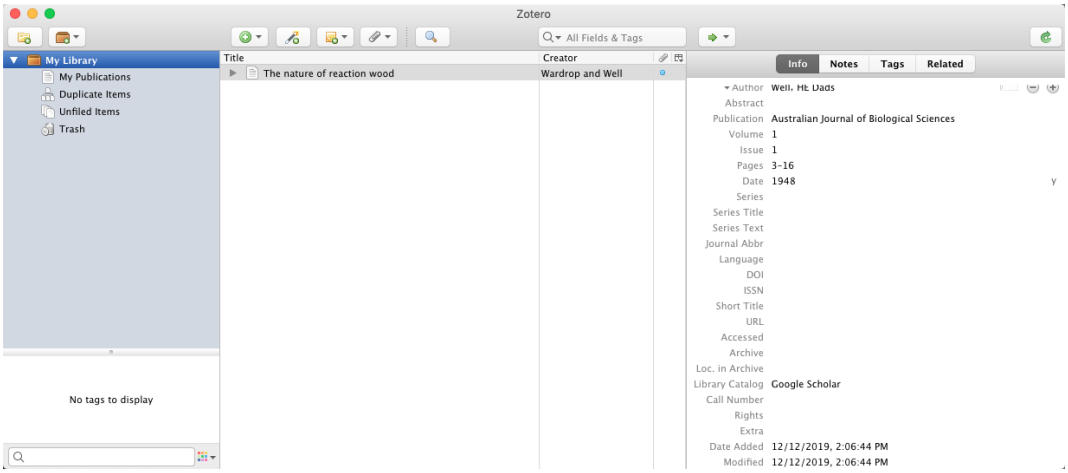


Figure 2: A screenshot of Zotero

The Zotero has the most efficient web browser plug-in. With the help of Google Scholar, the Zotero can achieve very good efficiency in creating new entries of reference. Users also have personal accounts for sync and managing references. Like Mendeley, Zotero also helps in acquiring the PDF files of references.

2.3.3 EndNote

EndNote is a commercial, closed-source reference management application by Clarivate Analytics, which also runs the Web of Science. The EndNote has its own format of library users can use to import/export. It also has a "database search" integration. Though the EndNote does not offer a browser integration, the reference entry can be downloaded as .enw file and added to the library. A screenshot of the EndNote's interface is shown in Figure 3.

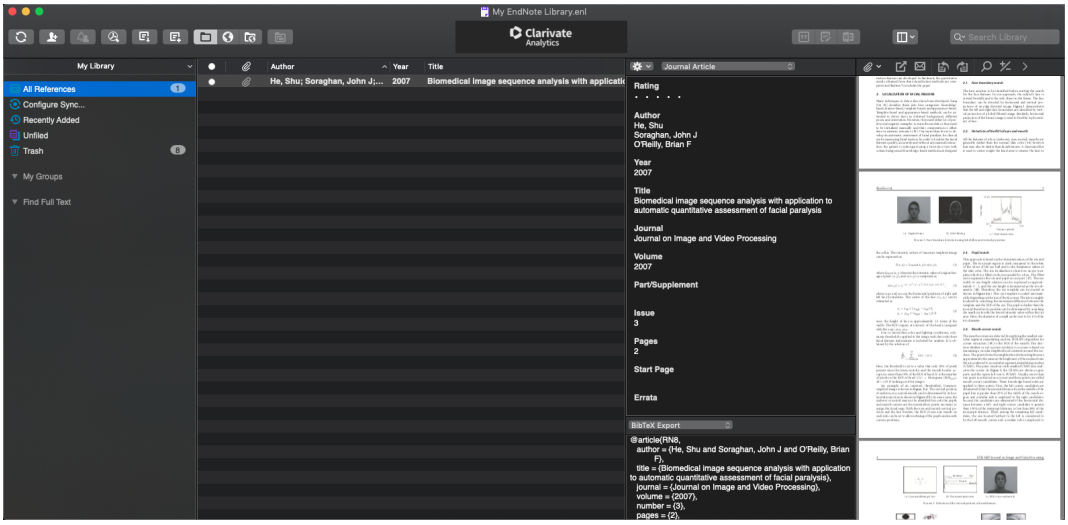


Figure 3: A screenshot of EndNote

The EndNote requires the purchase of subscription, which substantially prevents a number of users from using it. It has an integration with Google Scholar so that the reference can be imported from Google Scholar in 1-click.

2.3.4 RefWorks

RefWorks is a web-based reference management application produced by ProQuest. RefWorks requires an institutional subscription. RefWorks relies on the institutional verification of each personal account. It also has a Google Scholar integration. A screenshot of the RefWorks' interface is shown in Figure 4.

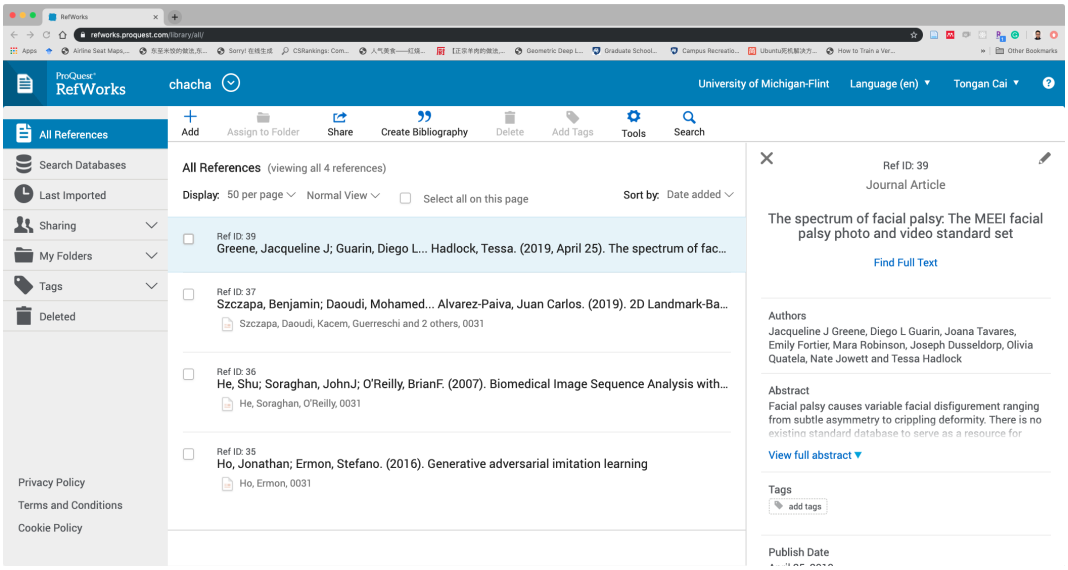


Figure 4: A screenshot of RefWorks

The RefWorks allows a strong auto-completion with only the title information of the paper. Another virtue of RefWorks is that RefWorks does not require any installation, which is good for working on multiple devices or public computers. Also, the BibTeX content export of RefWorks can be directly copied to clipboard, which is very efficient for LaTeXusers.

2.4 Validation of RUI

The RUI is a software to record keystrokes and mouse behaviors Kukreja, Stevenson, and Ritter (2006). It generates results as text files, indicating the timestamp of each mouse movements and keystrokes. Researchers applied RUI in studying robot teleoperation tasks (Ritter, Kukreja, & Amant, 2007), learning and retention (Kim, Koubek, & Ritter, 2007) and so on, and the software is considered accurate in analyzing mouse movements, clicks, and keystrokes. A screenshot of the RUI running on Windows is shown in Figure 5.

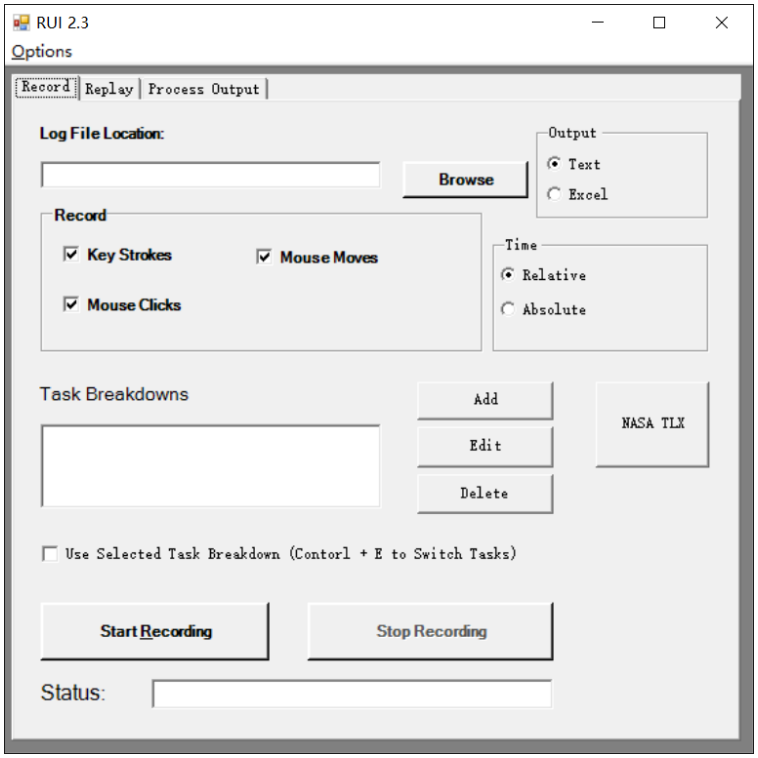


Figure 5: A screenshot of RUI

We would first demonstrate that RUI is accurate in recording user behavior before conducting controlled experiments. We will prove that 1) the RUI can accurately record the movements of mouse point and keystrokes, 2) the RUI maintains the same accuracy across different operating systems, 3) Whether there's lag or disorder between recording of keystrokes and mouse movement.

We first examine that the RUI can accurately capture the motion of the mouse. On both Windows and OS X, a PowerPoint slide with a letter-sized canvas is opened. The verification

procedure goes as:

1. Scale the canvas to 100% (actual size)
2. Turn on RUI and start recording
3. Hold cursor at one coordinate, record the coordinate shown in Paint and left-click mouse
4. Repeat 3. multiple times
5. Stop recording on RUI
6. Get the log file, look for "Pressed Left" and the coordinates those entries give.
7. Compare the relative position of each coordinate recorded manually and from log file

We tested 10 points and all of them give a pixel-wise accurate recording. We believe that the RUI is extracting mouse information from a low level of system.

The examine of accuracy for keystrokes is trivial, we simply pressed keys from A to Z while RUI is recording and the log file successfully detected all of them. To examine the potential lag or disorder between keystroke and mouse movement or click, we quickly and repeatedly left-click the mouse, followed by pressing a key in the keyboard. The RUI successfully captured all the behaviors in the correct order. Then we kept moving the mouse, while pressed a key for 5 times. The RUI collected exactly the number of keystrokes. We think the RUI may collect mouse and keyboard behaviors in separate scripts.

It can be concluded that RUI is accurate in recording mouse behaviors and keystrokes on both Windows and OS X operating systems. For the log files generated by RUI, we leave out the headers and extracted the information section, saved as a text file so that we can do counts of keystrokes and calculate the total mouse movements by simple Python programming.

2.5 Design & Procedure

The study can be separated into three stages. In the first stage, a pre-study survey is conducted on the usage and usability issues of reference management applications. In the second stage, quantitative study will be piloted on subjects measuring their physical and mental effort across different applications. In the third stage, one retrospective survey will be given and the subjects in the second stage are asked again about their opinions towards these reference management applications.

2.5.1 Pre-study Survey

In the first stage, the survey questionnaire first ask about the general usage of reference management applications. The survey starts by asking the participants if they use the reference management applications and their frequency of use. Then, it asks the participants to pick or give their favorite reference management application. The survey then suggests the usability issue of reference management applications and ask if the users are aware of how

long it can take to add a new entry to their collection, whether they think these applications are not efficient enough, as well as the factors they think that determines the overall quality of a reference management application. The participants who finished the survey can volunteer in the quantitative study in the second stage.

2.5.2 Quantitative Study

In the second stage, the required tasks flow like the following. For each of the reference management application on each of the operating systems,

1. The subject is given a list of titles, without elaborated author or publisher information. (S)he is required to give a full bibliography management list with accurate information and the corresponding PDF files.

2. The subject will first acquire the PDF files of the papers from Google Scholar. During this process, (S)he can use the Chrome Plug-in or Google Scholar integration to import the file into the reference management applications.

3. The subject then works on the reference manager's auto-complete function by only fill in the title of the paper. (This method work for Mendeley & RefWorks)

4. If no auto-complete is available, then the subject will try to open an "Online search" inside the reference manager, which has a direct link to online databases. (This feature is provided by EndNote)

5. If neither of the features are available, the subject will go back to Google Scholar and obtain the information of the reference.

6. After obtaining the references, the subject is also responsible to check that the information provided by the reference manager/online database/Google Scholar is accurate. If not, she will have to make corresponding changes for the entries.

7. The subject will then clean up and format the entries, including date format/author name/publisher, etc.

8. The subject then output a bibliography list of papers in APA style, together with a .bib file including all the references.

Each subject is asked to follow the required task and operate the applications on the two operating systems while being recorded by the RUI. For calculation of mouse movement, we simply adopted a sum of Euclidean distances between each recorded state that are labeled as "Move". Also, the total number of mouse clicks are presented. We try to eliminate bias by using a keyboard to navigate between Chrome and reference manager applications.

The Cogulator is then used to formulate a separate Keystroke-Level Model model for each application. The Keystroke-Level Model (KLM) a variation of GOMS that was proposed by Card, Moran, and Newell (1980). KLM has a set of physical and mental operators,

including keystrokes, button clicks, keyboard to mouse movement, pointer movement, as well as thinking time to predict how long it will take expert users to complete a specific task with no errors. Each application will require a different set of operators in different execution orders, which will result in different ideal mental processing time for the users that will be modeled by the Cogulator, and the built KLMs are verified and agreed by the participants to ensure no operation was neglected.

2.5.3 Retrospective Survey

The retrospective survey aims to evaluate the subjects' change of opinion after trying different reference management applications in a relatively short time period and find out what are the important factors for them to choose his/her favorite reference management application.

3 Experiment & Results

3.1 Results of Survey on Reference Manager

Of the 24 participants of the pre-study survey, 11 of them report that they use reference management applications. The proportion of reference management usage is rather low among graduate students. One possible reason is that more than half of the participants are first-year master students, who may not be very heavily involved in academic research.

Among the 11 participants who use reference management applications, 3 of them report everyday usage of the applications, 4 of them will use the applications several times a week. 2 report that they use the reference management applications less than once a week but more a couple of times in a month. The other 2 participants use the applications less than once a month. The result indicates that participants do have some experience with the reference management applications and can at least represent a group of general users.

The question of favorite reference management application shows that 5 out of 11 participants prefer to use Mendeley. There are 3 who like Zotero, 2 EndNote and 1 RefWorks. The majority of the participants choose to use a free version of such applications (RefWorks requires subscription).

The survey then asks how long it will take to add one new citation to your reference manager. 6 out of 11 assert that this will be under 1 minute, 3 think it will take 1 to 2 minutes and 2 regard this to be between 2 minutes to 5 minutes. We infer that many of the Mendeley and Zotero users simply uses the web browser plug-in to add the new entries but did not consider the time to acquire the PDF file and correct all the information on the entry.

7 of 11 participants believe that the efficiency of the current reference manager is low and should be improved, 2 participants who report less than 1 minute for the previous question do not think there is a need to improve the efficiency. 2 other think the efficiency does not

matter.

The extra question asked about the participants' evaluation of whether a reference management application is good. The result is shown in Figure 6.

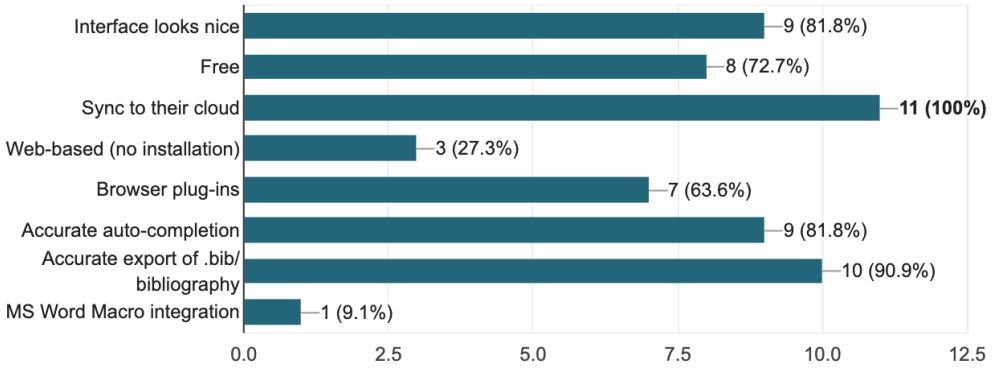


Figure 6: Summary of answers for evaluation of reference management applications.

All participants believe that sync to cloud storage is important. Popular virtues include accurate bibliography export, nice interface, accurate auto-completion and free of charge.

Summarizing the pre-study survey, we can see that for the majority of users of reference management application users, the efficiency still matters, yet users tend to overlook the process of cleaning/formatting and acquiring PDF files.

3.2 Functionalities & Features

As is shown in Table 1, we listed and summarized all the important features in the nowadays commonly used reference managers. We investigate 5 main categories of the functionalities: (1) Import file; (2) Database connectivity; (3) Complete and correct basic information; (4) Export bibliography; (5) Operating system support.

All of the reference managers show complete support for import file formats. As for Database Connectivity, Zotero can connect to more databases than the other three. Both Mendeley and Zotero are better than the other two in terms of basic reference information completion and correctness. When exporting bibliography, all of the reference managers provide various formats. Last but not least, all of the four applications support Windows and OS X. However, only Mendeley and Zotero support Linus. As for mobile applications, Mendeley and EndNote provide iOS App. Only Mendeley provides an app for Android.

In conclusion, among all of them, RefWorks has the most number of features. However, there is no simple conclusion as to which reference manager is the best, in terms of functionalities and features. To suit users' different situations, users may choose different reference managers.

Task	Mendeley	Zotero	EndNote	RefWorks
Import File				
Manually Add	•	•	•	•
Add Entry from PDF	•	•	•	•
Import from .bib/.RIS etc	•	•	•	•
*Browser Plug-in	•	•	•	•
Database Connectivity				
ArXiv	•	•	•	
CiteSeer	•	•	•	
IEEE Xplore		•	•	
PubMed	•	•	•	•
Complete & Correct Basic Information				
Manually Change	•	•	•	•
Online Detail Filling	•			•
PDF information extraction	•			•
Export Bibliography				
Different formats	•	•	•	•
Export .bib	•	•	•	•
*MS Word Integration	•	•	•	•
Operating System Support				
Windows	•	•	•	•
MacOS	•	•	•	•
Linux	•	•		
iOS App	•		•	
Android App	•			

Table 1: Comparison of the functionalities and features among applications

3.3 Task Analysis with Cogulator

The Cogulator is an open-source application for human performance calculation. One screenshot of the Cogulator's interface is shown in Figure 7. The Cogulator provides a series of basic operators for Task Analysis, and it takes a natural language syntax. Each of the operators has a default execution time, and the Cogulator allows users to apply different integration of the operators. The output of the Cogulator is quantified as time for the task to be completed.

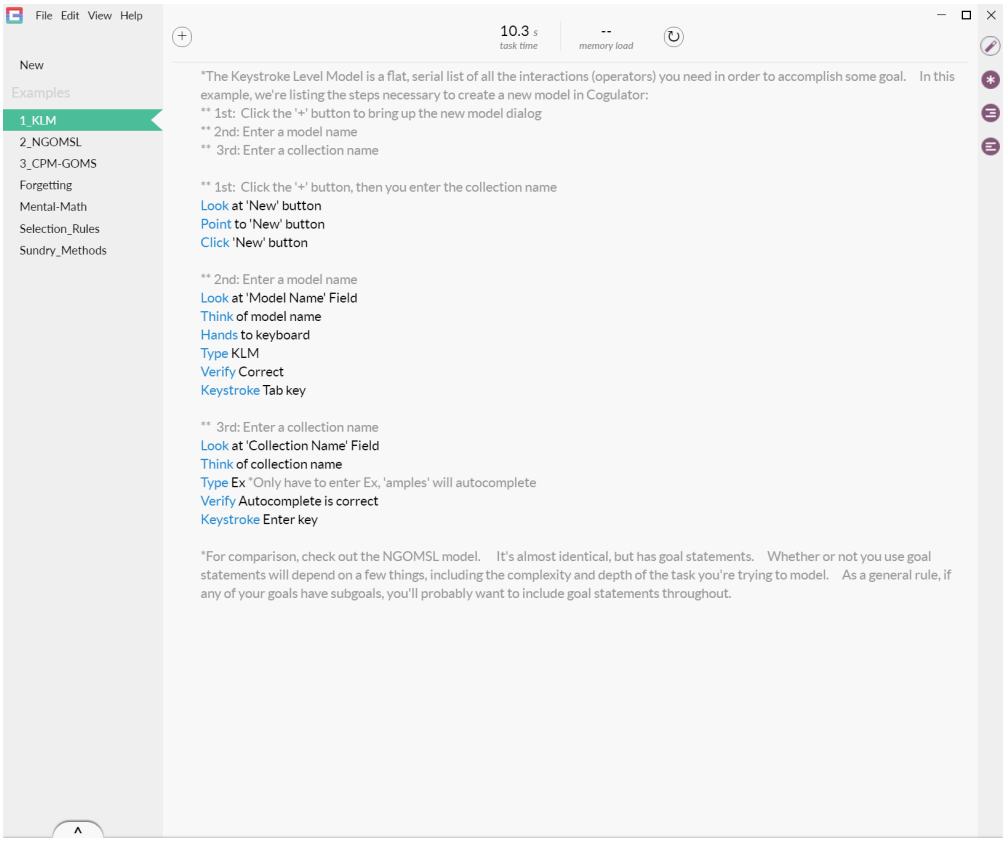
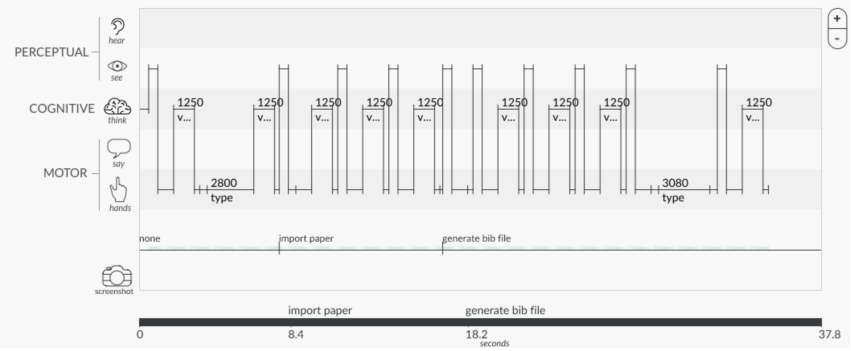


Figure 7: A screenshot of the Cogulator's interface.

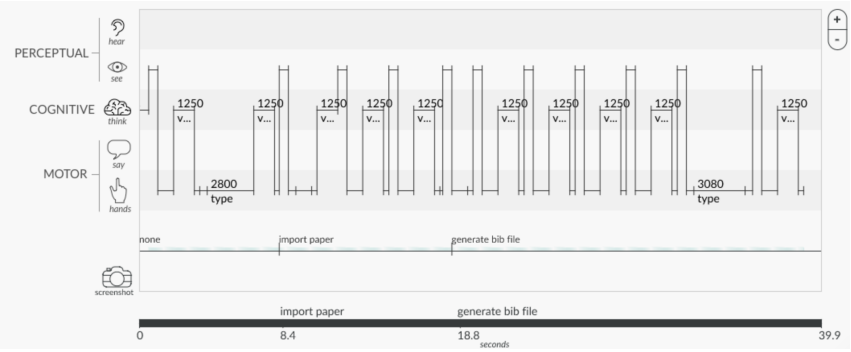
Specifically for KLM, there are operators including click/keystroke (type for multiple keystrokes)/point and also look, think, verify. The operators adopted in this study are (the time are given by the default values in Cogulator):

- Click, mouse operation, 0.3s
- Point, mouse movement, 1s
- Keystroke, keyboard, 0.3s
- Type, continuous keystrokes, time varies
- Hands, physical movement of hand, 0.5s
- Look, cognitive process, 0.6s
- Think, cognitive process, 1.3s
- Verify, cognitive process, 1.3s

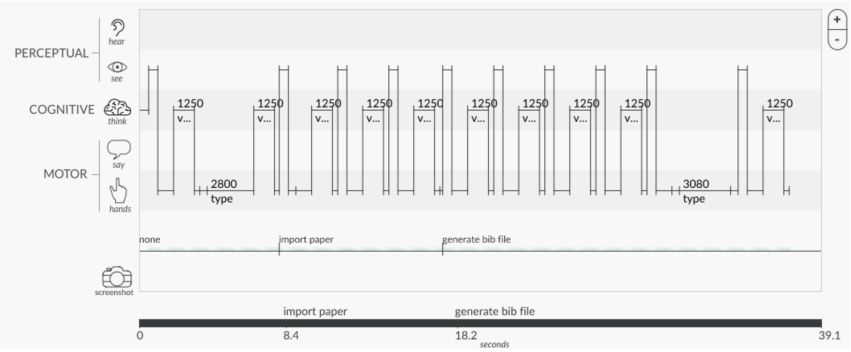
Results and discussions As shown in table 2, RefWorks achieves the most efficient performance in completing the task, with the least total execution time and number of operators involved. To better illustrate the discrepancy between each reference manager, the graph of load (Gantt Chart) on each aspect of the user is shown in Figure 8. The detailed scripts of Cogulator are attached in the Appendix section.



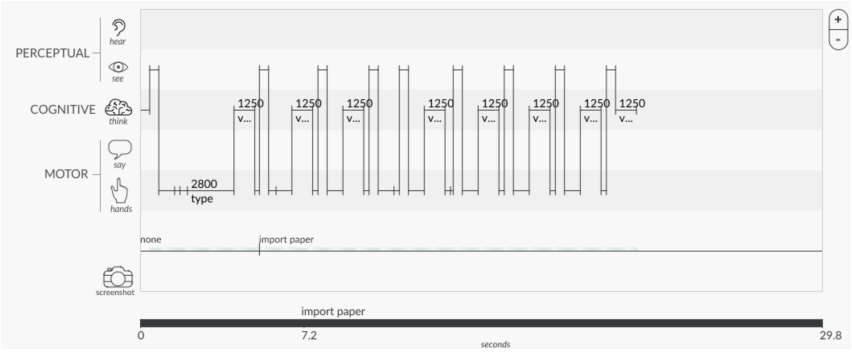
(1) Mendeley



(2) Zotero



(3) EndNote



(4) RefWorks

Figure 8: Graphic presentation of Cogulator results

	Total Execution time (s)	Number of Steps	Number of Memory chunks
Mendeley	37.8	50	1
Zotero	39.9	53	1
EndNote	39.1	51	1
RefWorks	29.8	41	1

Table 2: The results of the cogulator for different reference managers

3.4 Statistics Reported by RUI

We tested Mendeley, Zotero, EndNote, and RefWorks on both Windows and Mac OS. The task is to import 3 selected papers into the reference manager software and the bulk export biblatex formatted bibliography file. We evaluate the efficiency of these tools in terms of the time elapsed, distances of mouse movements and the number of mouse clicks. The distances and time is rounded to the nearest integer.

As is shown in Table 3 and Table 4, in terms of the average number, Zotero is better than all the other reference managers in all aspects. Zotero shows significantly better performance, both in Windows and OS X. However, when we look at the individual numbers, results may vary.

Besides Zotero, Mendeley, as the most popular free reference manager software [5], shows good performance in both of the operating systems. Specifically, in terms of all the three aspects, Subject#1 in MacOS shows its effectiveness in Mendeley. Also for Subject#4 in Windows, Mendeley performs relatively well.

EndNote costs a lot in MacOS but it is the best in Subject #1, #2 in Windows in terms of the time elapsed. One of the reasons could be that google scholar support downloading a .ris formatted file which could be used to directly import into the EndNote.

RefWorks can be a safe choice among all the others and for all the subjects. The performance of RefWorks would not likely fall into the worst. However, RefWorks rarely performs the best in all scenarios.

[5] According to the G2 reference managers review in 2019.
https://www.g2.com/categories/reference-management

Mouse distance (Pixels)						
Application \ Subject	#1	#2	#3	#4	#5	Average
EndNote	76812	73916	49751	84166	99324	76794
Zotero	**58173**	**64310**	**45455**	**49073**	**77763**	**58955**
Mendeley	90628	77637	59191	57304	143965	85745
RefWorks	65191	91058	63591	83624	117360	84165
Number of mouse clicks						
Application \ Subject	#1	#2	#3	#4	#5	Average
EndNote	103	77	73	95	182	106
Zotero	**94**	91	**61**	65	**81**	**78.4**
Mendeley	**94**	78	70	**61**	141	88.8
RefWorks	77	**74**	63	86	94	78.8
Elapsed Time passed (Seconds)						
Application \ Subject	#1	#2	#3	#4	#5	Average
EndNote	**162**	**150**	98	219	196	165
Zotero	168	177	**80**	**158**	**143**	**145**
Mendeley	290	221	167	249	255	236
RefWorks	180	202	116	235	185	184

Table 3: Results reported in Windows

Mouse distance (Pixels)						
Application \ Subject	#1	#2	#3	#4	#5	Average
EndNote	62263	71374	99532	74512	118725	85281
Zotero	50394	65194	**43394**	**41047**	**72364**	**54479**
Mendeley	**28691**	107756	64050	45975	73934	64081
RefWorks	52430	**61743**	76652	58091	100787	69940
Number of mouse clicks						
Application \ Subject	#1	#2	#3	#4	#5	Average
EndNote	103	98	113	70	113	99.4
Zotero	67	62	**49**	54	**47**	**55.8**
Mendeley	**44**	74	59	**51**	60	57.6
RefWorks	67	**55**	65	54	60	60.2
Elapsed time passed (Seconds)						
Application \ Subject	#1	#2	#3	#4	#5	Average
EndNote	208	142	322	113	196	196
Zotero	152	103	**141**	**74**	**115**	**117**
Mendeley	**115**	144	229	134	192	163
RefWorks	161	**100**	167	92	148	134

Table 4: Results reported in Mac OS X

3.5 Results of Retrospective Reports

The retrospective survey aims to collect feedback from the 5 subjects who just participated in the quantitative study. After they finished the experiment, each of them received the link to the survey to guarantee the feedback can be consistent with their short-term memory.

The survey first asks the participants to quantitatively rate their experience with each of them. On a 1-5 standard, Mendeley received an average of 3.8/5; the Zotero got amazingly 5/5; the EndNote got 2.6/5, and the RefWorks got 4.2/5.

The ratings coincide with the question "Which one is the best reference management application?" as 4/5 participants rated the Zotero as the best one and one rated the RefWorks. The only one who rated RefWorks as favorite is probably very used to the application, as we noticed in the following question if the participants changed their minds. 3 of them chose yes since they now have Zotero as their preference.

We also asked about the estimated time it will take to add one new citation reference manager. This time, 3 of them agree that it can be less than 1 minute, and the rest 2 thinks it can be achieved within 1-2 minutes. We also asked which application takes the longest time to operate. The reported result varies, 3 saying EndNote and one each for RefWorks and Mendeley. This could be due to subjects' different levels of familiarity with the applications.

We asked again how will the participants tell if a reference manager is good. The result is shown in Figure 9.

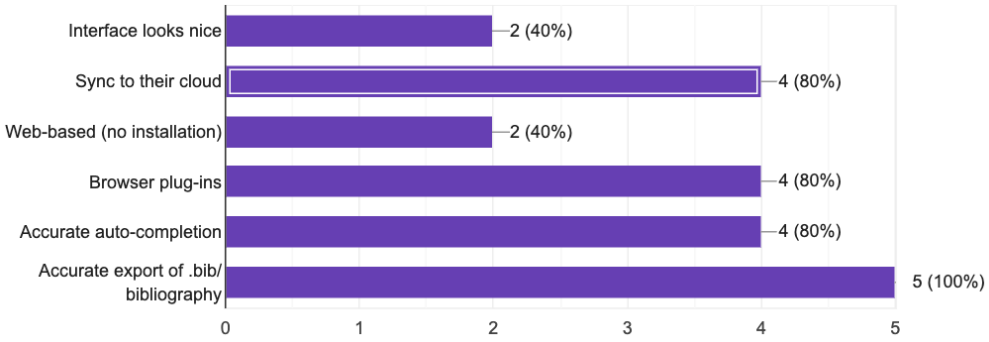


Figure 9: Retrospective answers for evaluation of reference management applications.

Comparing to the previous result, users seem less focused on the interface quality, but more on performance metrics like accurate export of references.

The retrospective shows that our study does raise the users' awareness of the usability, efficiency and accuracy issues of the reference managers. The preference for reference management applications can be related to the application's performance. Most users may not get exposed to all types of reference management applications and tends to report the first application they are familiar with as their preference.

4 Conclusion & Discussion

Reference management applications are widely used in academia and research community to support the query and management of references, which is of great convenience in the development of manuscripts. This project tries to develop a way of evaluating the quality of reference management applications quantitatively and try to find the connection between the users' physical and mental efforts in using these applications to the users' preference for such applications.

The results of RUI and retrospective surveys show that users tend to prefer the application with the least physical effort to operate (Zotero) for the specific task. Comparing the retrospective preference to their answers in the pre-study survey, we can also conclude that the variation in user preferences can result from unfamiliarity with other reference management applications. As reported by the participants, those who changed their minds have not got a chance to try Zotero or RefWorks. A very large portion of users are already comfortable with the current preference that achieves exactly the same goal. As one of the participants asserted before the study, "I have been using EndNote for may years. It's not free but convenient to use." After trying between the four applications, he suggested that "Zotero and RefWorks are so good! I'd probably stop my EndNote subscription." It can be inferred that if users got a chance to compare the applications on the same task, they might reach a different conclusion. Promotion of reference management applications should let users to physically try it, either by providing tutorials for institutional subscribers or online sessions for novices—instead of giving out advertisements around the internet.

From the feature table and task analysis results, it can be concluded that these four reference management applications have almost the same design of features and functionalities, but may in different forms of presentations. Some of them (Mendeley, Zotero) may have browser plug-ins and others could use Google Scholar integrations (RefWorks, EndNote). The feature itself may not able to directly cause a discrepancy in user experience, but the efficiency, responding time and correctness matters. The Mendeley and Zotero can both import a reference entry using the plug-in and acquire the PDF files at the same time. However, all participants experienced a long wait time when using Mendeley. The plug-in for Mendeley runs for a long time before providing the users with a list of papers shown on the webpage. Zotero, instead, finishes in seconds. The integrations of EndNote and RefWorks do not help to acquire the PDF files for the entries, which will add to the efforts users have to make if they want entries with PDF. The presentations of features can be an important factor that the developers should consider.

This work stems from the previous qualitative reviews of reference management applications and further considers the quantitative mental/physical efforts of users. The proposed method improves the traditional two-stage method (pre-study survey plus participants study) by introducing a mental simulation and retrospective survey. The idea of evaluating the user's mental effort and change of thoughts for a specific task is not widely adopted in the comparative study of applications, and we believe this method can reveal closer to real feedback from users (participants) regarding their experience with the applications.

We also recognize the limitations of this work. For the group of subjects, although there is some diversity in their habits of using reference management applications, the group can not well represent the actual users of the applications. Senior graduate students, research professors, librarians can be expected to use such applications with a much heavier workload, for which their participation may lead to different study results. The measurement of the physical effort of users is based on a small simple task, which may not be comprehensive in revealing the applications' full functionalities or all of the users' needs. Full task analysis can be conducted to provide more insights into the applications' usability. Future may also adopt the proposed method in analyzing other applications or transfer the idea to other user-oriented analyses.

References

Al-Badareen, A. B., Selamat, M. H., Jabar, M. A., Din, J., & Turaev, S. (2011). Software quality models: A comparative study. In *International Conference on Software Engineering and Computer Systems* (pp. 46–55). Springer.

Basak, S. K. (2014). A Comparison of Researcher's Reference Management Software: Refworks, Mendeley, and EndNote. *Journal of economics and behavioral studies*, *6*(7), 561–568. Retrieved from http://www.econis.eu/PPNSET?PPN=817025995

Bellamy, R., John, B., & Kogan, S. (May 21, 2011). Deploying CogTool: Integrating Quantitative Usability Assessment into Real-world Software Development. (pp. 691–700). ICSE '11. ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1985890

Card, S. K., Moran, T. P., & Newell, A. (1980). The Keystroke-level Model for User Performance Time with Interactive Systems. *Commun. ACM*, *23*(7), 396–410. doi:10.1145/358886.358895

De Cristofaro, E., Du, H., Freudiger, J., & Norcie, G. (2013). A Comparative Usability Study of Two-Factor Authentication. Retrieved from https://arxiv.org/abs/1309.5344

Gilmour, R., & Cobus Kuo, L. (2011). Reference Management Software: A Comparative Analysis of Four Products. *Issues in Science and Technology Librarianship*, *66*.

Kim, J. W., Koubek, R., & Ritter, F. (2007). Investigation of procedural skills degradation from different modalities. In *Proceedings of the 8th International Conference on Cognitive Modeling*, Ann Arbor, Michigan, USA.

Kukreja, U., Stevenson, W. E., & Ritter, F. E. (2006). RUI: Recording User Input from Interfaces under Windows and Mac OS X. *Behavior Research Methods*, *38*(4), 656–659. doi:10.3758/BF03193898

Lorenzetti, D., & Ghali, W. (2013). Reference Management Software for Systematic Reviews and Meta-analyses: An Exploration of Usage and Usability. *BMC medical research methodology*, *13*, 141. doi:10.1186/1471-2288-13-141

Mead, T. L., & Berryman, D. R. (2010). Reference and PDF-Manager Software: Complexities, Support and Workflow. *Medical Reference Services Quarterly*, *29*(4), 388–393. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/02763869.2010.518928

Park, K. S., & Hwan Lim, C. (1999). A Structured Methodology for Comparative Evaluation of User Interface Designs Using Usability Criteria and Measures. *International Journal of Industrial Ergonomics*, *23*(5), 379–389. Retrieved from https://www.sciencedirect.com/science/article/pii/S0169814197000590

Ritter, F., Kukreja, U., & Amant, R. (2007). Including a Model of Visual Processing With a Cognitive Architecture to Model a Simple Teleoperation Task. *Journal of Cognitive Engineering and Decision Making*, *1*(2), 121–147. doi:10.1518/155534307X232811

Santharooban, S., & Lavakumaran, L. (2018). A comparative Analysis of Reference Managers: EndNote and Mendeley. In *Proceedings of the 2018 International Conference of University Librarians Association of Sri Lanka* (pp. 96–107).

Appendix

Python script for RUI statistics

We attach the Python script to process the RUI records as below.

```
import pandas as pd
from scipy.spatial import distance
import os

file_list = os.listdir("./Logs")

for file in file_list:

    if 'win' in file:
        data = pd.read_csv(os.path.join('Logs',file),sep='\t')
        moved_data = data[data['Action']=='Moved']
        moved_data.X = moved_data.X.astype('float64')
        moved_data.Y = moved_data.Y.astype('float64')
        distance_list = []
        for i in range(moved_data.shape[0]-1):
            coordinate = (moved_data.iloc[i]['X'],moved_data.iloc[i]['Y'])
            coordinate2 =  (moved_data.iloc[i+1]['X'],moved_data.iloc[i+1]['Y'
                ])
            distance_list.append(distance.euclidean(coordinate, coordinate2))
            # calculate every two rows euclidean distance

        distance_list.append(0)

        moved_data['euclidean_distance'] = distance_list

        print("Distance sum: ",moved_data['euclidean_distance'].sum())
        print("Number of mouse clicks:",data[data['Action'].str.contains("
            Pressed")]['X'].count())
        print("Elapsed Time passed: ", data['Elapsed Time'].max() - data['
            Elapsed Time'].min())
        print(file)
```

Listing 1: Python processing script

Mendeley KLM Model

Goal search for the paper
Recall paper name
Look search bar
Point search input bar
Verify cursor is over the target
Click the input bar
Hands to keyboard
Type paper name
Verify correct
Keystroke enter

Goal import paper
. Look at paper title
. Hands to mouse
. Point to paper title
. Verify cursor is over the target
. Click paper
. Look at side bar
. Point web plug-in
. Verify cursor is over the target
. Click web plug-in
. Look choices
. Point to target
. Verify cursor is over the target
. Click import

Goal generate bib file
Swipe windows back to Zotero app *previously look up is in web browser
.Look sidebar
.Point All Document
.Click All Document
.Look for papers
.Point papers needed
.Verify cursor is over the target
.Click the papers
.Look tools bar
.Point the file in leftside upper corner
.Verify cursor is over the target
.Click file
.Look export
.Point export
.Verify cursor is over the target
.Click export
.Look the prompted window
.Point save as
.Hands to the keyboard
.Type bibfilename
.Hands to mouse
.Look save
.Point save
.Verify cursor is over the target
.Click save

Zotero KLM Model

Goal search for the paper
Recall paper name
Look search bar
Point search input bar
Verify cursor is over the target
Click the input bar
Hands to keyboard
Type paper name
Verify correct
Keystroke enter

Goal import paper
. Look at paper title
. Hands to mouse
. Point to paper title
. Click paper
. Verify cursor is over the target
. Look at side bar
. Point web plug-in
. Verify cursor is over the target
. Click web plug-in
. Look choices
. Point to target
. Verify cursor is over the target
. Click import

Goal generate bib file
Swipe windows back to Zotero app *previously look up is in web browser
Look sidebar
.Point MyGroup
.Click MyGroup
.Look for papers
.Point papers needed
.Verify cursor is over the target
.Click the papers
.Look tools bar
.Point the file in leftside upper corner
.Verify cursor is over the target
.Click file
.Look export library
.Point export library
.Verify cursor is over the target
.Click export library
.Look the prompted window
.Point ok
.Verify the export format
.Click ok
.Look the input bar
.Hands to the keyboard
.Type bibfilename
.Hands to mouse
.Look save
.Point save
.Verify cursor is over the target
.Click save

EndNote KLM Model

Goal search for the paper
Recall paper name
Look search bar
Point search input bar
Verify cursor is over the target
Click the input bar
Hands to keyboard
Type paper name
Verify correct
Keystroke enter

Goal import paper
. Look at paper title
. Hands to mouse
. Point to paper title
. Verify cursor is over the target
. Click paper
. Look at EndNote text
. Point EndNote text
. Verify cursor is over the target
. Click EndNote text
. Look Downloads
. Point the target file
. Verify cursor is over the target
. Click the target file

Goal generate bib file
Swipe windows back to Zotero app *previously look up is in web browser
.Look sidebar
.Point MyGroup
.Verify cursor is over the target
.Click MyGroup
.Look for papers
.Point papers needed
.Verify cursor is over the target
.Click the papers
.Look tools bar
.Point the file in leftside upper corner
.Verify cursor is over the target
.Click file
.Look export
.Point export
.Verify cursor is over the target
.Click export
.Look the prompted window
.Point save as
.Hands to the keyboard
.Type bibfilename
.Hands to mouse
.Look save
.Point save
.Verify cursor is over the target
.Click save

RefWorks KLM Model

Goal search for the paper
Recall paper name
Look search bar
Point search input bar
Click the input bar
Hands to keyboard
Type paper name
Verify correct
Keystroke enter

Goal import paper
. Look at paper title
. Hands to mouse
. Point to paper title
. Verify cursor is over the target
. Click paper
. Look at RefWork text
. Point RefWork text
. Verify cursor is over the target
. Click RefWork text
. Look Continue button at the new prompted web
. Point the Continue button
. Click the Continue
. Look Okay button
. Point the Okay button
. Verify cursor is over the target
. Click the Okay button

Swipe windows back to RefWorks web *previously look up is in web browser
.Look sidebar
.Point MyGroup
.Verify cursor is over the target
.Click MyGroup
.Look for papers
.Point papers needed
.Verify cursor is over the target
.Click the papers
.Look tools bar
.Point export button
.Verify cursor is over the target
.Click export
.Look the prompted texts
.Verify the prompted texts are the correct bib texts

Survey on Reference Manager Usage

Please provide your answer to the following questions. Thanks!

* Required

1. **Do you use reference managers?** *
   Mark only one oval.
   - ◯ Yes
   - ◯ No *After the last question in this section, stop filling out this form.*

2. **If so, what reference manager do you prefer to use?**
   Mark only one oval.
   - ◯ Mendeley
   - ◯ EndNote
   - ◯ Zotero
   - ◯ RefWorks
   - ◯ EasyBib
   - ◯ Other: _____

3. **How often do you use your preferred reference manager?**
   Mark only one oval.
   - ◯ Everyday
   - ◯ Several times a week
   - ◯ Less than once a week
   - ◯ Less than once a month

4. **How long it will take approximately for you to add one new citation to your reference manager?**
   Mark only one oval.
   - ◯ Less than 1min
   - ◯ 1min-2min
   - ◯ 2min-5min
   - ◯ >5min
   - ◯ I have no idea

5. **Do you believe that the efficiency of current reference manager is low and should be improved?**
   Mark only one oval.
   - ◯ Yes
   - ◯ No
   - ◯ Does not matter

6. **How will you tell if a reference manager is good? Select all apply**
   Check all that apply.

   - [ ] Interface looks nice
   - [ ] Free
   - [ ] Sync to their cloud
   - [ ] Web-based (no installation)
   - [ ] Browser plug-ins
   - [ ] Accurate auto-completion
   - [ ] Accurate export of .bib/bibliography
   - [ ] MS Word Macro integration
   - [ ] Other: _____

7. **Any Suggestions?**

_____

_____

_____

_____

_____

Ask for participation! (optional)

We're conducting a study on evaluating the efficiency of reference managers and we're looking for participants. The study will ask you to operate several reference managers on Windows and OS X for some simple tasks. If you're willing to participate in this study please leave your email here.

8. **Email:**

_____

Retrospective Survey

Please rate your experience with Mendeley:

	1	2	3	4	5	
Very Bad	○	○	○	○	○	Very Good

Please rate your experience with Zotero:

	1	2	3	4	5	
Very Bad	○	○	○	○	○	Very Good

Please rate your experience with EndNote:

	1	2	3	4	5	
Very Bad	○	○	○	○	○	Very Good

Please rate your experience with RefWorks:

	1	2	3	4	5	
Very Bad	○	○	○	○	○	Very Good

Which one is the best reference management application?

- ○ Mendeley
- ○ Zotero
- ○ EndNote
- ○ RefWorks

Did you changed your mind?

- ○ Yes
- ○ No

How long it will take approximately for you to add one new citation to your reference manager?

- ○ Less than 1min
- ○ 1min-2min
- ○ 2min-5min
- ○ >5min
- ○ I have no idea

Which application takes the longest time to operate?

- ○ Mendeley
- ○ Zotero
- ○ EndNote
- ○ RefWorks
- ○ I have no idea

How will you tell if a reference manager is good? Select all apply

- ☐ Interface looks nice
- ☐ Sync to their cloud
- ☐ Web-based (no installation)
- ☐ Browser plug-ins
- ☐ Accurate auto-completion
- ☐ Accurate export of .bib/bibliography
- ☐ Other: _____

Any suggestions?

_____