

Usability Evaluation of Mendeley Literature Search Feature



March 1, 2019

1 Introduction

Researchers from various backgrounds and of different levels of expertise in searching for literature require tools to find and add references to their database (Mead and Berryman, 2010). In the present study, we investigate the literature search feature of the reference manager *Mendeley* in its Desktop version for macOS. Here, we do not focus on systematic literature searches (which may require specific keyword searches, or construction of citation graphs (Hart, 2001)), but investigate the simpler case of needing to find a reference in the software given a reference in another form. More specifically, a user may be writing a scientific paper and, remembering the name, year and some part of the title of that publication, wants to find the full reference in order to add it to her library. Here, depending on individual search strategies employed, users may prefer to search e.g. according to author and year or the title ¹.

In the Mendeley Desktop application, search for references that are not already in one's personal library is implemented as a simple search bar, as presented in Figure 1A. The search is designed such that a user may type in her query with specific tags in the form of e.g. `author:[author] year:[year] title:[title]`. Crucially, Mendeley requires this information to be filtered using the specific tags that can either be typed in or inserted by clicking on the magnifying glass icon and selecting the correct field, as shown in Figure 1B. In order to use this feature, users would either have to know about

¹Note that due to the implementation of the Mendeley search feature, searching for the reference by copying the full reference does not yield the correct result.

the tag system, or recognise that clicking on the magnifying glass provides the option to insert tags. However, given the size of the icon users may not find this feature, and falsely assume that the search functionality is the same as familiar search engines, such as *Google scholar*, where no such keywords are necessary and the input is instead interpreted more flexibly.

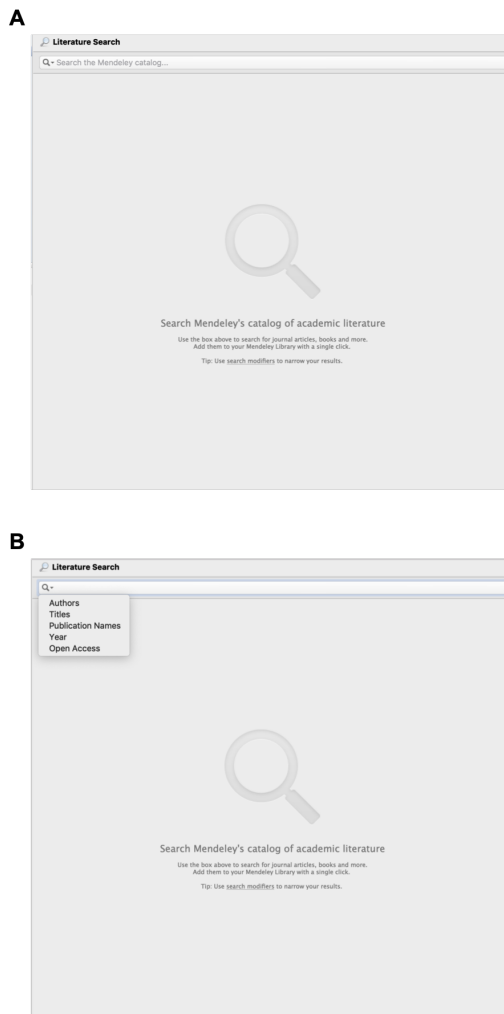


Figure 1: User interface for Mendeley reference search. **A)** Initial search field. **B)** Indication of search tags when clicking on the magnifying glass symbol

Here, we study a small sample of users in a task where three references

of different popularity need to be found. The scenario is that a user needs to find and add the corresponding references to his library. We measure the effectiveness of a trial, where success is operationalised as the user finding the correct reference, efficiency measured as task time in seconds and user's views on the system measured with the System Usability Scale (Brooke et al., 1996). Furthermore, search strategies were observed and qualitative comments recorded.

2 Method

$N = 4$ participants were recruited from the course *The Human Factor* at the University of Edinburgh, where participation in the study was voluntary. All participants were familiar with reference managers and 75% indicated that they use Mendeley as their primary reference manager. 50% of participants were female and 50% male. The average age was 23.25 years ($SD = 1.26$).

Prior to completing any tasks, participants were informed about the study (Appendix A) and gave their informed consent (Appendix B). In order to ensure consistent task instructions across participants, participants were given an instruction sheet (Appendix C) that described the tasks, where the investigator was available to answer any questions.

The task comprised finding the full references for three scientific articles, which were provided as complete references in APA style. When the participant indicated her readiness, the search was timed using a stopwatch, the search strategy observed by the investigator and all comments by the participant written down by the investigator. After a task was finished based on the participant clicking on "save reference", the time was recorded, the reference was checked for correctness, after which the timer was started again for the next task. In order to minimise ordering effect, the references were presented in a random order between participants.

After all tasks were finished, participants were given the System Usability Scale (Brooke et al., 1996) and could indicate any further comments in an extra text field (see Appendix D for the implementation of the questionnaire, which was administered in a paper-pencil format). Finally, participants were debriefed about the Mendeley search feature.

Due to the small sample size, the data were only analysed descriptively. Qualitative data were provided by participant's comments and from the text field at the end of the questionnaire.

Table 1: Task times were computed by taking the average of all tasks. Measures of dispersion for task time were computed across all subjects and tasks. Accuracy is computed by taking the proportion of successful tasks for each subjects, where the values presented here provide estimates of the distribution of accuracies across subjects.

	Mean	Median	SD	IQR	Min	Max
Task Time [sec]	71.50	48.50	52.04	57.00	23.00	169.00
SUS Total Score	35.62	32.50	10.08	9.37	27.5	50.00
Accuracy [%]	75.00	83.33	31.91	41.67	33	100

3 Results

Measures of effusiveness, efficiency, and general usability are presented in Table 1. As visual inspection of the data indicated skewness particularly for task times, robust measures of central tendency and dispersion are also reported. Comments and textual responses at the end of the survey are presented in Table 2.

On average, participants evaluated the system unfavourably, as indicated by a mean SUS score of 35.62, which falls in the *not acceptable* range, according to Bangor et al. (2008). Task times were around one minute, and searches successful 75% of the time. Here, participants sometimes found partially correct references, where some entries would have to be edited manually.

Observing search strategies revealed that no participant made use of the tag system or clicked on the magnifying glass symbol.

4 Discussion

The present study highlights the difficulty in understanding the search feature of Mendeley, where user’s behaviour and comments indicated that the system is misaligned with user’s mental models. The present results show negative effects on effectiveness, efficiency and subjective experience.

We see two main approaches for possible solutions. One would be to provide a search mask that explicitly asks for information of the corresponding type. Such a solution is outlined in Figure 2, which depicts the PubMed search interface. This solution would not change the underlying search im-

Table 2: Qualitative comments made by participants during and after the experiment.

Comments made during the experiment	Comments made after the study
“Why is there no advanced search feature here?”	“I use Mendeley almost every single day and I never knew there were tags for searching.”
“This is why I never search in Mendeley and go to Google scholar instead.”	“I learned to search for titles in Mendeley - author and year simply don’t work. ” “Literally searching for the full reference didn’t give me anything useful, which I found confusing.” “Even now that I know about the feature, I don’t think I will ever use the tool and search elsewhere or add the reference manually.”

plementation, but only the interface by making the (tagged) information that is required more explicit to the user. Even though previous research has shown that overall, only a small proportion of users make use of advanced search features such as Boolean operators (Spink et al., 2001), this may be different for users of reference managers, as this user group is more likely to be educated in effective search strategies than the general population.

Figure 2: User interface for PubMed advanced search with exemplary selection of author and publication date fields. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/advanced>

Instead of requiring the user to learn and memorise specific tags for searching, we suggest to adopt a system-centric view (Reason, 2000) and make the interface more intuitive and explicit, in order to reduce errors and improve task times.

Another (but technically more challenging) solution would be to adapt the search engine back-end, such that information entered by users into one search field would be interpreted, similar to standard general purpose search engines. Ideally, users could choose between an intuitive interpreted search field and a more systematic search mask with different categories, as provided e.g. by PubMed.

References

Bangor, A., Kortum, P. T., and Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human-Computer*

- Interaction*, 24(6):574–594.
- Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.
- Hart, C. (2001). *Doing a literature search: a comprehensive guide for the social sciences*. Sage.
- Mead, T. L. and Berryman, D. R. (2010). Reference and pdf-manager software: complexities, support and workflow. *Medical Reference Services Quarterly*, 29(4):388–393.
- Reason, J. (2000). Human error: models and management. *Bmj*, 320(7237):768–770.
- Spink, A., Wolfram, D., Jansen, M. B., and Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American society for information science and technology*, 52(3):226–234.

A Participant Information Sheet

Participant Information Sheet

What this study is about

The purpose of this study is to understand how people use the reference manager *Mendeley*. The study is conducted as part of a coursework requirement for the course *The Human Factor* at the University of Edinburgh.

Your participation in this study is voluntary

You can take a break at any time. Just tell the researcher if you need a break. You can leave at any time without giving a reason.

Information we want to collect

We will ask you to show us how you use the product. We will watch and record how you do various tasks and we will ask you some questions.

How we ensure your privacy

Any personal information will only be stored and processed locally on an encrypted and password protected computer. We may share reports that include your comments and actions but your data will be anonymous. This means your name and identity will not be linked in our research reports to anything you say or do.

Your consent

Please sign this form showing that you consent to us observing your during the experiment and collecting the data described above. If you want to withdraw your consent in the future, contact the person named below who will destroy any personal data we hold about you. Otherwise, we will delete your personal data after two years.

**Investigator contact details
redacted**

Signature: _____

B Consent Form

Consent Form – Usability Study

I agree to participate in the study conducted by [REDACTED] as part of a coursework requirement for the course *The Human Factor* at the University of Edinburgh.

I understand that participation in this usability study is voluntary and I agree to immediately raise any concerns or areas of discomfort during the session with the study administrator.

Please sign below to indicate that you have read and you understand the information on this form and that any questions you might have about the session have been answered.

Date: _____

Please print your name: _____

Please sign your name: _____

Thank you!

We appreciate your participation.

C Participant Instructions

Instructions

Imagine you are currently writing a literature review and need to find a number of reference to add to your reference manager. You are given the reference and should try to find the full reference in Mendeley and add it to your library. Try to use the search strategy that you would typically use to find references, i.e. you can use as much information from the given reference as you think is adequate to find the given entry in Mendeley.

Task 1

Please find the following reference and save it to your library:

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4), 297-323.

Task 2

Please find the following reference and save it to your library:

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.

Task 3

Please find the following reference and save it to your library:

Buzsaki, G. (2002). Theta oscillations in the hippocampus. *Neuron*, 33(3), 325-340.

D System Usability Scale

Mendeley Desktop - Literature Search Evaluation

For each of the following statements, mark one box that best describes your reactions to the Mendeley search tool today. There are no wrong answers.

1. **I think that I would like to use this system frequently.**

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly Agree

2. **I found the system unnecessarily complex.**

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly Agree

3. **I thought the system was easy to use.**

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly Agree

4. **I think that I would need the support of a technical person to be able to use this system.**

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly Agree

5. **I found the various functions in this system were well integrated.**

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly Agree

6. **I thought there was too much inconsistency in this system.**

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly Agree

7. I would imagine that most people would learn to use this system very quickly.

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly Agree

8. I found the system very cumbersome to use.

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly Agree

9. I felt very confident using the system.

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly Agree

10. I needed to learn a lot of things before I could get going with this system.

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly Agree

Comments

11. Do you have any other concerns or comments that you would like to add?

Thank you for your participation!
