

Research Exchange - A Collaborative Paper Annotation Tool

Kalpesh Krishna
kalpesh@cs.umass.edu

Yueying Liu
yueyingliu@cs.umass.edu

Shashank Mehrotra
smehrotra@umass.edu

Jiazheng Zhu
jiazhengzhu@umass.edu

ABSTRACT

This research looks to explore the requirements, development, and efficacy of the online research paper tool-Research Exchange. Based on surveying graduate and undergraduate research students, it was found that getting started with research is a huge undertaking and requires constant guidance of immediate peers and professors. To address these issues, a collaborative online research tool was designed to help young researchers understand concepts on the fly. To understand whether the tool helps achieving the desired goal, usability studies were conducted to understand the usability and functionality issues of the tool. It was found that collaboration using annotations was extremely beneficial for users which were in line with our findings from the requirement phase. Additionally, it was noted that users appreciated the online reader and were open to trying it from the existing pdf solutions. However, many users noted several additional features that would be necessary to make this product useful. Future work involves improving the prototype and integration of the OpenReview APIs for large scale deployment. Our final prototype can be found on <http://researchexchange.cs.umass.edu:3000/>.

INTRODUCTION

Reading academic paper has never been easy, especially for new researchers or researchers needing to read a paper from a different field. To understand the problems with paper reading faced by new researchers, we conducted 12 interviews with new researchers with different profiles; we also obtained 3 screen recordings of graduate students reading academic papers using their preferred way on desktops. There are many challenges that new researchers face when starting out to read papers.

- The technical jargon are hard to understand and the online explanations sometimes are not suitable for the context of the specific paper.
- There are many online resources that help with paper understanding. However they are scattered and each individual researcher needs to repeat the effort of finding them, resulting in a lot of waste in time and effort.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CS690A Final Report

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN N/A.

DOI: [N/A](#)

- Evaluating the relevance of a paper usually takes a skim through the paper as the abstracts are sometimes not enough.
- Paper formatting are inconsistent across conferences, journals and fields. Even for experienced researchers, reading papers from a new field takes some time to adapt.

As a result, reading papers requires a lot of institutional support to help new researcher make the transition. There is heavy reliance on the research community around them which may not be accessible to them at all times. While there are plenty of referencing tools and information retrieval systems available for use, there is a lack of tools that can enable new researchers to read papers in context of their budding needs. Therefore we seek to alleviate these issues from a human-centered-design point of view.

The objectives of this project has two folds. On the individual researcher level, we intend to develop a collaborative research paper reader that can aid researchers to understand the papers better; on the academic community level, we want to enable aggregation of annotations which would help researchers gain different perspectives.

We created Research Exchange that allows annotations on both document level and sectional level. The document level view only contains the abstract and metadata of a paper; the sectional level view breaks the paper into sections for easy navigation and allows the researchers to read the paper by sections. In the document level view, we have annotations for the entire paper for easy determination on whether the paper is worth reading. In the sectional level view, parts of the text from a section can be highlighted and have annotations associated with. There are three types of annotations for both views to solve different problems.

1. **Supplementary Materials** contain links to external resources that help with the understanding of the paper. The links can be directed to blogs, videos, other papers, etc.
2. **Comments** contain general information about the entire paper or a specific line. This can be explanations for technical jargon, point out small mistakes in the paper, provide more details of the paper in the document level view, etc.
3. **Questions** allow researchers to post questions about the entire paper or some specific part of a paper if the other annotations cannot answer the questions. The other researchers, including the author can answer these questions.

We conducted user tests with 10 researchers using our product and compared it with the researchers' previous way of paper reading. The results showed that they like the features of Research Exchange and are willing to use it, although some

new features need to be added. Our main findings from the analysis were:

- Annotations significantly aid user understanding, especially for inexperienced researchers.
- Section level annotations are more helpful than document level annotations.
- Researchers are willing to share their responses in a research group. Researchers are probably more willing to share results within their research group rather than publicly.

LITERATURE REVIEW

Paper Reading

The advent of low cost digital screens has encouraged screen reading of academic papers. When screen reading was introduced in 1990s, survey conducted by O'Hara and Sellen [11] found several shortcomings with screen reading methods including poor annotation tools, poor navigation and document layout. While several advancement were made in screen reading, a survey by Liu [8] found limited adoption of "in-depth" reading and annotation of documents via electronic media. Majority of these concerns have been addressed in modern PDF softwares with better annotation tools. Walsh [15] found minimal differences in a reader's speed and comprehension ability between print and screen formats and noticed that existing PDF readers do not use features which do not have counterparts on paper. The survey recommended the design of interactive multimedia features to enhance paper reading experiences. Recent work has tried to use digital methods in paper reading via QR codes [5] and noticed a significant improvement in reading comprehension via "scaffold questions" - pre-reading questions designed to help readers understand what to expect from the document, and post-reading questions designed to test a reader's understanding of the document.

The paper reading experience varies greatly across people in different stages of their academic careers [7], with young researchers often struggling with the methods sections and older researchers undervalue the critical analysis of data, unlike senior researchers. Mentorship from senior researchers (advisors) plays an important role in training young researchers to be skilled readers [13].

Collaborative Reading

Previous research shows that conversational collaborative learning environments help not only the receiver but also the giver [16, 2]. Social annotation tools have been shown to be effective in enhancing the learning experience [10]. More concretely NB¹, a pedagogical tool used to collaboratively annotate course material [17, 18] has gained widespread popularity. Other products similar to our project are nurture.ai and <https://web.hypothes.is/>. Our main points of difference are the segregation of the document and sectional view as well as fine-grained annotation types (Comments, Questions and Supplementary Material).

Paper Relevance

Keywords are the most popular methods to find relevant papers, often using TF-IDF for better retrieval [12]. People also look

at citations to judge paper relevance HITS and PageRank of citation graphs are popular algorithms to build reading lists of relevant papers [6]. Semantic Scholar² uses automatic techniques to convert scientific literature into a structured representation [1, 14], the *Literature Graph* which facilitates easy querying of scientific literature [3]. It also uses citation information to recommend relevant papers [4]. Prior work has also shown a correlation between readership (via Mendeley) and citation count [9], making readership count a rough metric for citation count.

NEED FINDING

Participants

We chose a total of 12 participants for this study. We took notes for the initial 8 unstructured interviews, and fully transcribed the final 4 detailed structured interviews - the ones that were conducted in a more detailed and focused manner. In addition, we procured 3 screen recordings of people reading papers. We chose a diverse set of participants with the following criteria -

1. **Experience** - 3 participants were senior undergraduate students at UMass, who were just starting research. The other participants were more experienced readers.
2. **Industry vs Academia** - 2 participants were currently working in the industry. The other participants were in an academic setting (UMass).
3. **Diverse Backgrounds** - We chose a diverse set of backgrounds for our interviewees - computer science, electrical engineering, industrial engineering, linguistics and biochemistry.

To cater our initial points of focus, we chose participants who were

1. Different stages of paper writing
2. Non-native speakers of English
3. Occasionally reading papers outside their area of expertise to aid their research.

Location of the Interviews

Interviews were conducted at public spaces (close to where the interviewee usually work), offices and laboratories. A few clarification questions were asked on via social media (Slack, Messenger). Screen recordings were conducted by participants in isolation (as per their convenience) - without the interference of any interviewer.

Interviews

We started the study with preliminary observations first. We used the 8 unstructured interviews with a set of diverse focal points to understand the experience of reading academic papers by researchers. These interviews lasted between 20-30 minutes.

Each interviewer took notes during the interviews on points which they found interesting. The team members had a meeting to share notes and findings about the initial interviews. We then decided on the above focal points by discussing among team members on aspects like the universality of the problem,

¹nb.mit.edu

²<http://semanticscholar.org/>

personal feelings, availability of target groups around us, importance of the issue in a study, etc. These focal points are not the same as the initial 8 focal points, but are generated through the initial interviews.

After the 3 final focal points were decided, each team member picked focal points to have focused interviews. Each one of the team members tried to avoid taking focal points similar to the initial focal points generated by the member to avoid assumptions. Each team member conducted a semi-structured interview with the assigned focal point. These interviews lasted between 15-20 minutes.

The data captured was in the form of notes, screen recordings and audio recordings of the interviewees. The audio recordings were transcribed by the team members in the format suggested in class. Software Temi was used by some members to assist transcribing.

In addition, we also collected 3 screen records by asking researchers to record 15 minutes of the screen when they do their research. We did this to have some reference data other than the interviews.

Identified Focal Points

Based on the observation studies and the interviews, we narrowed down to three focal points. The process of arriving at the focal points involved the development of affinity diagrams and finding common themes based on the information obtained from the interviews. While this process was detailed, it is beyond the scope of the current document. We shortlisted to the following focal points:

1. Search for related papers during revision
2. Experts reading papers outside their domain
3. Interactive reading of papers

Findings around Chosen Focal Points

The researcher's experiences of reading papers in the process of submitting/ revising or framing their research problem

We found that during the process of submissions, the process of retrieving relevant papers can be difficult owing to the lack of relevant papers being recommended that could be included. Currently, researchers either rely on community (online/offline) support or look to references of the existing papers. It is difficult to read papers from a certain limited focus and thus ends up being a long process due to non-availability of contextual search.

Experts reading papers outside their area of expertise

This was one of the key pain points we found in our findings. Due to the nature of cross disciplinary research, there is heavy reliance on expert reviews to understand research papers. Additionally, due to unfamiliarity with the domain, there is a lot of time spent trying to determine the right keywords.

How interactive reading of papers impact the user experience of the researcher

We found that exposure to "E-formats" and availability of complementary multimedia, blog posts, subtext to equations and other material helps in the understanding the paper better. It is however an aid and the process of reading papers is something all researchers are heavily entrenched in. The interactive

readings aid the researchers in active note-taking, compiling and focusing on their research better.

All of these findings are explained in greater detail under different key-findings in the next section which could help put our analysis in perspective.

SYSTEM

Low Fidelity Prototype

We built a paper prototype as our low-fidelity prototype and used it to obtain initial feedback about our project. The paper prototype needed a facilitator for operation and supported some forms of interactivity like folded drop-down lists and movable form components. We used small sticky notes to simulate the highlighting experience. Figure 1 shows our section level view. A full presentation with our paper prototype can be found in <https://bit.ly/2PHBiEd>.

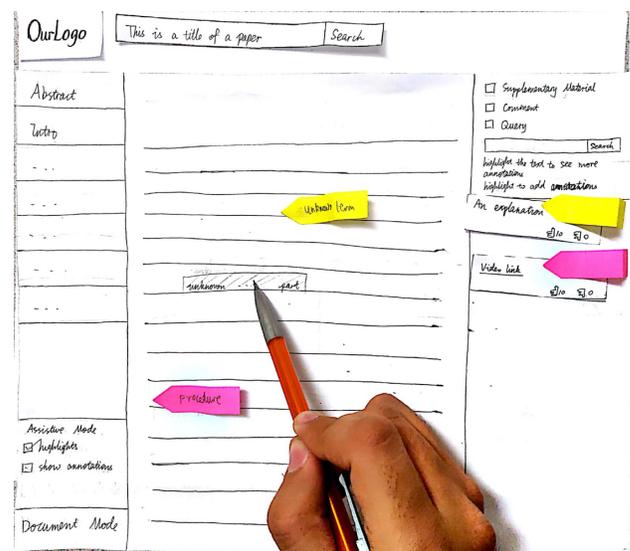


Figure 1. Section view of low-fidelity prototype

Medium Fidelity Prototype

Through the discussion of prototype design and user testing, we designed three initial medium fidelity prototypes interfaces. The following are screenshots and brief introductions of the three prototype interfaces.

Landing page

You could enter any matching titles in the search bar to search for research paper and the results will be sorted from top to bottom with relevance. Each of the searching results contains the title, authors, conference/journal and time of publication of a paper. (Figure 2)

Document Level Mode

You could find the abstract at the top of document mode. It helps you make a better understanding of the whole paper (there is also a Section Reading Mode button to enter the sectional mode). At the right side of the abstract text, you could see the authors, conference/journal, time of publication, DOI and everything related to the paper. In the "document level annotations" column, you can see the supplementary

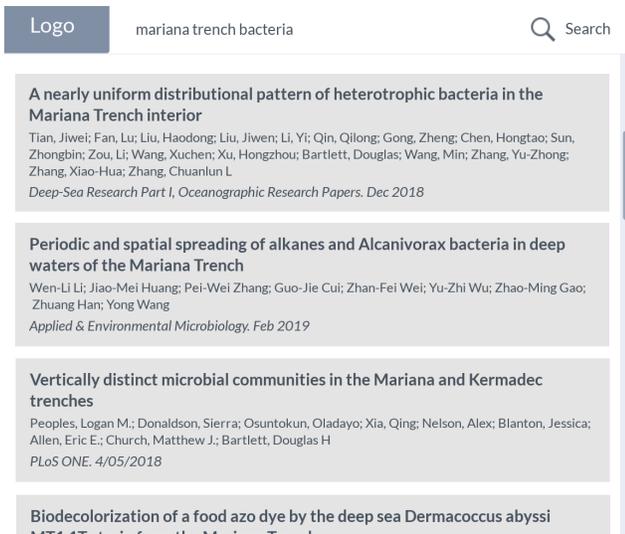


Figure 2. Landing View of medium-fi Prototype

materials, comments and queries. The checkboxes and search bar allows filtering through the annotations. (Figure 3)

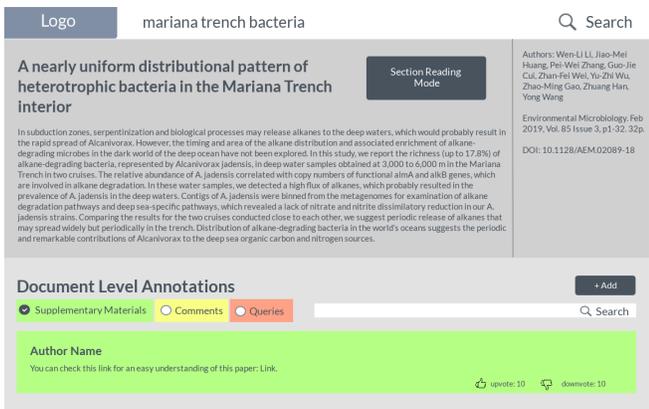


Figure 3. Document View of medium-fi Prototype

Section Mode

Finally, in our section level view we present a rough interface (without any of the highlights). We could not find an easy way to highlight text in Marvel. On the left bar, the sections allow readers to select specific sections whose full content to be displayed in the main area. The Assistant Mode allows the users to see or hide highlighting in text or annotations. On the right bar, users can filter and search the annotations made on specific sentences in the paper. The associated text of an annotation should be highlighted with the same color. (Figure 4)

Final Prototype

Overview of the whole prototype

This prototype works for researchers entering any new field to improve the reading efficiency. We designed the web application with three interfaces: landing/doc/section level. User can search the desired content through the landing page and then enter the document-wise level to see the

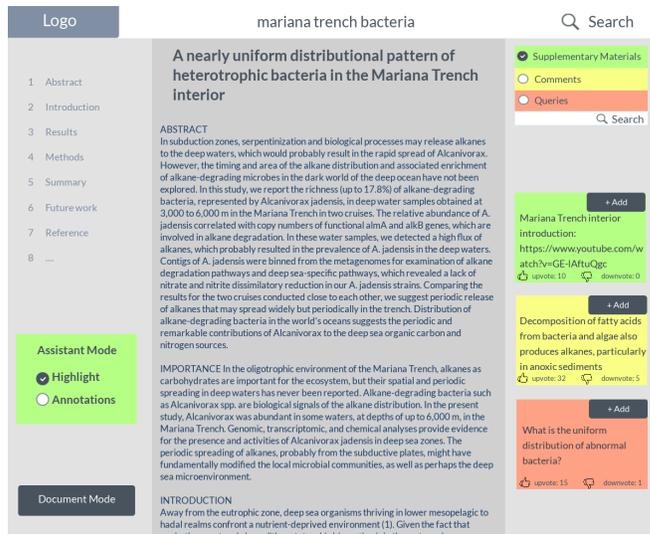


Figure 4. Section View of Low-fi Prototype

overview of paper. Users could also add annotations through the comment/question/supplementary interfaces. In the sectional level, people could read the paper section by section. In the process of reading, they could highlight the sentences/paragraphs according to their own will to add a comment/question/supplementary material.

Introduction

Based on experimental design and usability testing for medium fidelity prototype, we have made improvements and revisions, the following are the three prototype illustrations.

Landing page

Supported actions

- search for papers by keywords.

Unsupported actions

- login/logout.

Function description

The search bar is used to search the paper according to the keyword searching mechanism, It helps users to find the specific paper smoothly. Every auxiliary reading application needs the search bar to help people find the article they want. Our search interface is clean and simple, fully helping users to focus on the core appeal.

Results will be sorted from top to bottom with relevance. After the user clicked the search button, the system would sort a series of paper with respect to how their titles, authors and conference/journal match the keywords. The reader could click on the corresponding article to enter the doc-wise level to learn more detailed paper information. According to the arrangement of relevance, the reader can better find the articles they want to match. Generally, people can't remember the complete title when searching for articles, and this sort of search is appropriate for this situation. (Figure 5)

Document Level Mode

Supported actions

- browse abstract with document level annotations;

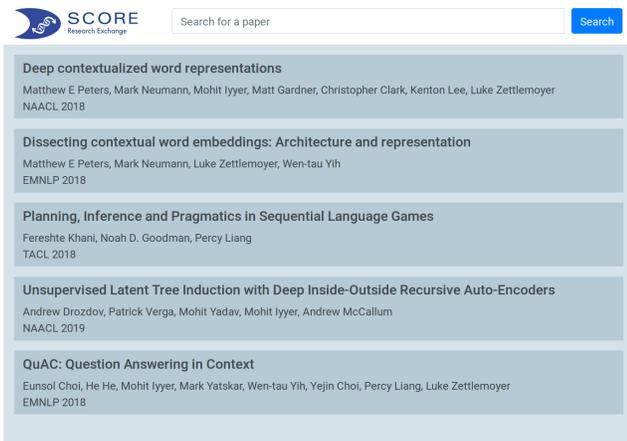


Figure 5. Landing View of Final Prototype

- add document level annotations;
- vote document level annotations.

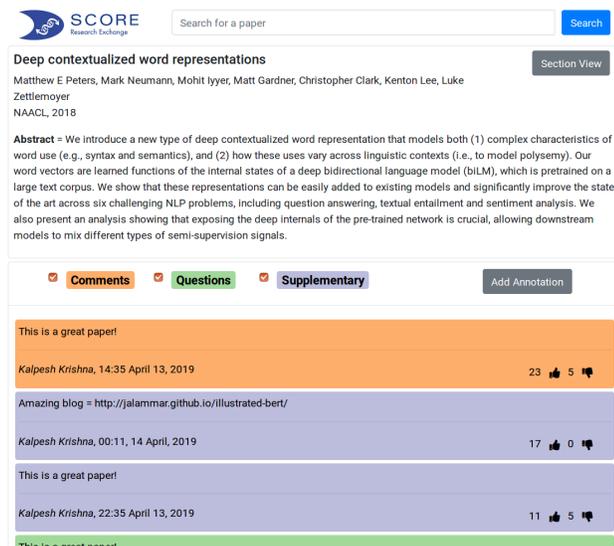


Figure 6. Document View of Final Prototype

Unsupported actions

- answer queries;
- only vote once;
- paper upload from user side;

Interface description

A reader could find the title, authors, publisher and abstract at the top of document view.

Below the abstract are annotations. This part is the same as the medium fidelity prototype except that we removed the search bar after receiving that due to the size of the research community, the number of annotations will be small and searching is not necessary most of the time.

If a user clicks on the add annotation button, a popup window with a text box and annotation type selection will be displayed. Users can add annotations themselves and the added annotation will be displayed later at the bottom of all the doc-level annotations.

If the paper is not desirable based on the information in this page, a reader could search for any other paper through the search bar at the top of the website. After searching, it will return to the previous interface. If the paper is useful after having an overview of the paper, the sectional view button in the upper right corner can be used to enter the sectional reading mode. (Figure 6)

Section mode

Supported actions

- Read paper sections / add sectional annotations;
- vote sectional annotations;
- hide/show highlights.

Unsupported actions

- answer queries;
- only vote once;
- add annotations with math equations/images.

Interface description

The sectional view is mostly similar to the medium fidelity prototype. We moved the toggle of highlights to the right so that all the checkboxes are together, and we have more space to accommodate papers with more sections. We also added the basic paper information on top for easy references. The doc-view button is moved to the same place as the sectional-view button so that it is easy to be found.

The complete paper content is in the middle of the interface. In the text, there are many sentences that are highlighted with different colors. These highlights shows that the text is related to annotations of supplementary materials, comments, and questions respectively, which are shown on the right side of the interface. By selecting certain paragraphs or sentences, the user can highlight and annotate the content of the paragraph (the specific operation is to hold down the right mouse button and select the text you want to comment on, then the add annotation button will appear at the top right of the interface. After clicking that button, the same popup window will appear and the operation is the same). (Figure 7)

EXPERIMENT

For the purposes of evaluation, we conducted the study in three phases. We wanted to emphasize realistic experimental settings. This is hard to emulate since the user's interface experience is critically dependent on the quality and quantity of annotations. In the context of our project we define a "realistic experimental setting" as one which people are naturally encouraged or expected to add annotations to parts of a research paper. We believe that we can find such settings in a university environment. With this in mind, we conduct experiments in three phases.

1. Research group readings
2. Structured interviews
3. large scale public peer-review via OpenReview

We believe that Phase I and Phase III are realistic experimental settings. Phase III would be the part of the future work for this research.

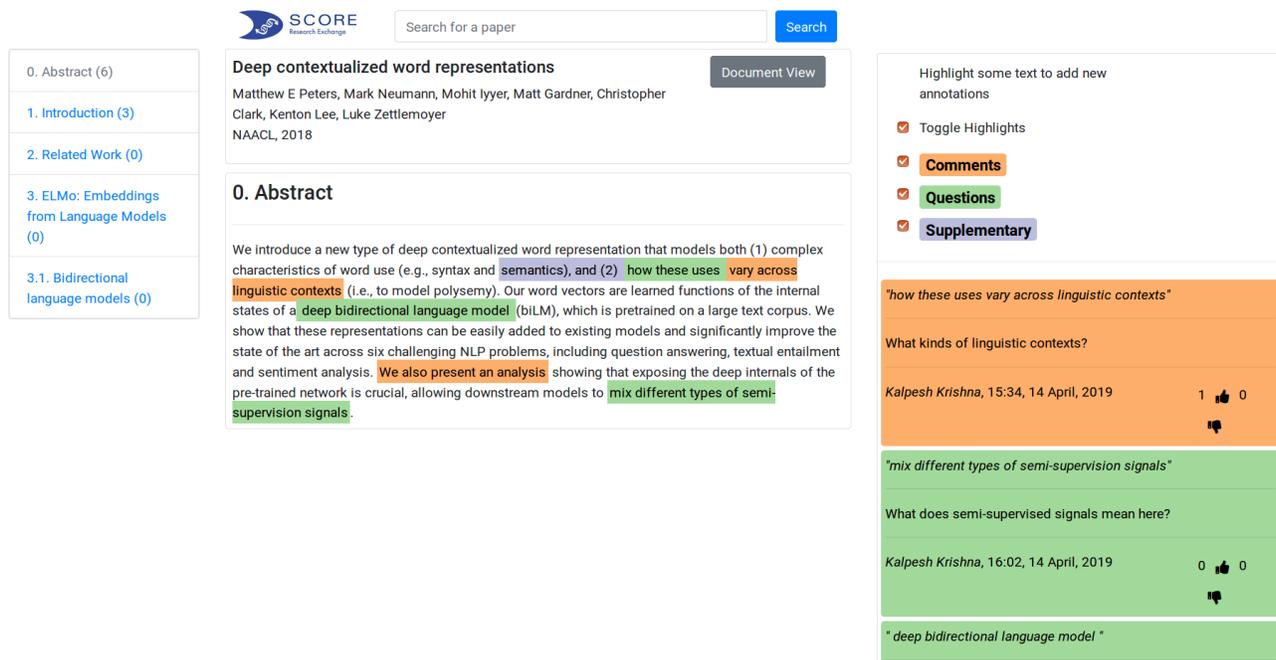


Figure 7. Section View of Final Prototype

Phase-I University Reading Group

A university reading group is a collection of students and faculty with a shared set of research interests who come together weekly and bi-weekly to discuss recent research papers or unpublished research by one of the participants. This encourages technical conversations among researchers in a university, encourages collaboration and helps beginners get more familiar with their field of interest. There is familiarity between the group members due to their constant interactions, collaborations and in some cases even mentor-mentee constructs between different group members. The reason for why this was chosen was that our tool was built for an audience of researchers with different research experiences. We assumed that a research group is representative of our target population and in a naturalistic setting, would be a great determinant of the validity of our design. The first phase involves the study of a research group of 7 people on UMass campus. They are guided by 1 professor and work and collaborate within the group. Their experience varies from experienced Ph.D. to sophomore students. The group is balanced by gender however in this study there were 4 males and 3 females. Additionally, there is an equal number of inexperienced and experienced students in the research group. Measures for the participants include their experience, the type of research they conduct (interdisciplinary vs within discipline) and their methods of searching, compiling and satisfaction with the current format of reading papers. In the second phase, three more people who are also research students but do not belong to any specific research group. Since we wanted to test on how intuitive is to use the tool, the task for the experiments were only to use the tool to assist with reading an academic paper. For the data collected until now, below table shows the summary.

Phase-II Reviewing the usability of the tool with individual researchers

While it is important to understand the usability of the group within a group, it is also important to evaluate the individual features by members who are not a part of the group. We conducted the review with 10 people (7 from the previous research group and 3 additional) for their analysis on the individual features and their usability. The distribution of participants has been shown in Table 1. It is a balanced distribution with a variety of expertise.

Phase-III Large Scale Peer-Review via OpenReview

OpenReview (<http://openreview.net/>) is a project initiated by the IESL laboratory in the Computer Science department at UMass Amherst. It is an attempt to change the messy peer-review process by making the whole process public. Any person can possibly add reviews or comments for papers. Our system's document level view is similar to the existing interface in OpenReview. While this is a long term goal of the project, we have initiated the first part of the conversations and are looking to integrate the existing Open Review APIs with our solution to make the application available to a larger user group.

RESULTS

Based on our experiments, we found qualitative and quantitative results from the two phases that we analyzed. We evaluated the interface in different contexts and found the efficacy of the interface. We found the emergent themes from the first phase which emphasized the features that were useful as well as features that could have been different. Additionally, the second phase helped us to determine how the individual sections of the tool added to the experience of the reader.

Phase	Experience	Description	Count
Phase I	Experienced Researcher	Completed PhD and > 4 years experience	1
	Research Student	PhD student with 2 to 4 years experience	3
	Beginner Researcher	Undergraduate student with 0 to 2 years experience	3
Phase II	Experienced Researcher	Completed PhD and > 4 years experience	1
	Research Student	PhD student with 2 to 4 years experience	3
	Beginner Researcher	Undergraduate student with 0 to 2 years experience	5

Table 1. Experimental results.

Qualitative Results

Based on the evaluation of the tool from the group, we found the emergent themes around the interface. The questions that were asked in the group discussion involved the discussion around the aspects of the tool they liked as well as those they wish could have been different. The discussion was conducted by one of the authors and compiled all the findings and based the themes on the comments.

What they liked

1. Appreciative of the different color coding of the different annotations within the tool
2. Like the dedicated space for asking questions within the application
3. Organization of the application and the layout is very useful

What they wished was different

1. Need for help and guidance feature for users
2. Separation between the paragraphs
3. Features like search; export of comments; private annotations would help with enhancing the usability of the tool

This result means that there is appreciation of the different features of annotations, the layout and the ability to ask questions is useful for the research group members. This is useful because it helps us understand that the translation of the requirements was inline with the expectations of users in a naturalistic setting. It is also important to note that while the application in its current form is enough to help users understand the tool, it is not enough for them to become active users due to rudimentary usability issues that need to be fixed. We plan to address these issues in the future.

Quantitative Results

We conducted the quantitative study of the application with 10 users, having a balanced distribution of experience (see Table 1). The objective was to understand the efficacy of the application based on survey questions that were asked to each user after they had used and evaluated the tool. All the responses were gathered using the Likert scale (1 to 5) with the options (1 - Completely Disagree, 2 - Somewhat Disagree, 3 - Neither Agree or Disagree, 4 - Somewhat Agree, 5 - Completely Agree). Based on these collated responses for each user, we tested our application on three parameters.

Effectiveness of Annotations

We first attempted to study whether having collaborative annotation in some format aids understanding. In a post-usage survey, we asked users "Did the annotations aid your understanding of the document?". We obtain the distribution of

responses shown in Figure 8. We first conduct a one-sample one-tailed t-test to verify the significance of this result.

H_0 : Researchers do not find any difference in their understanding by using the annotations. ($\bar{\mu} = 3$)

H_1 : **Annotations help improve researchers' understanding of papers** ($\bar{\mu} > 3$)

We obtained $\bar{\mu} = 4.4$, $t = 6.3317$, $df = 9$, and $p = 0.00005$ (which is less than $\alpha = 0.05$). Hence, this result is **extremely statistically significant**. Hence we reject H_0 .

As a baseline, we ask people "Are standard PDFs without any annotation sufficient in helping you understand the paper?". We obtain the results shown in Figure 9. Both "Somewhat disagree" responses were added by young inexperienced researchers. We use a one-tail two-sample t-test to test the following hypothesis,

H_0 : Researchers do not find any difference in their understanding between a raw PDF and an annotated PDF.

H_1 : **Annotations help improve researchers' understanding of papers over a unannotated PDF.**

We obtained $\bar{\mu}_1 = 3.6$, $\bar{\mu}_2 = 4.4$, $t = 2.1213$, $df = 18$ and $p = 0.0240$ (which is less than $\alpha = 0.05$). Hence, this result is **statistically significant**. Hence we reject H_0 .

Conclusion - Annotations significantly aid user understanding, especially for inexperienced researchers.

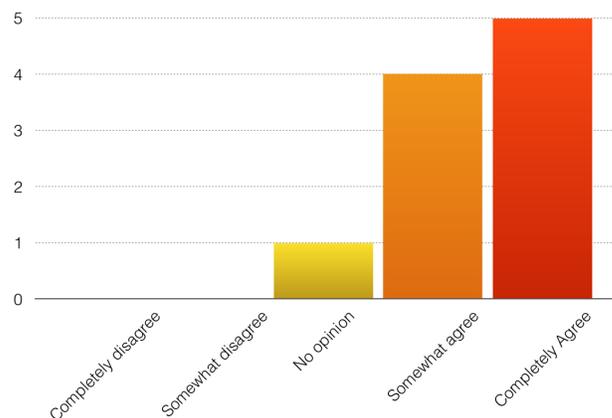


Figure 8. Are the shared annotations useful?

Section Level vs Document Level

We next tried to study whether fine-grained section level annotations are more useful than higher-level document level

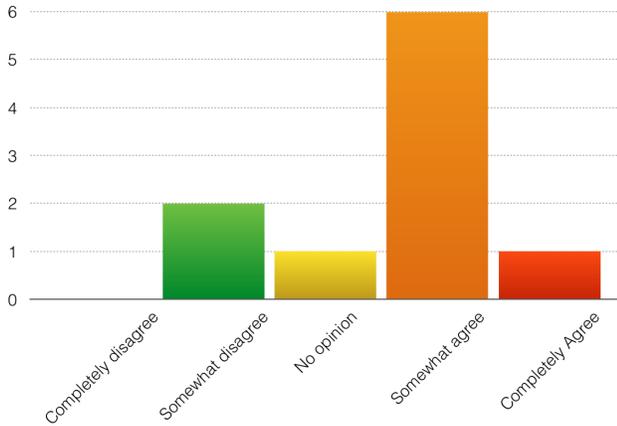


Figure 9. Do you feel unannotated PDFs are enough?

annotations. In a post usage questionnaire, we asked participants “Do you find section level annotations more useful than document level annotations?”. The distribution of results we obtain is shown in Figure 10. We conduct a one-tail one-sample t-test to verify statistical significance.

H_0 : Researchers do not find any difference in usefulness of document level and section level annotations. ($\bar{\mu} = 3$)

H_1 : **Annotations in the section level are more useful than annotations in the section level.** ($\bar{\mu} > 3$)

We obtained $\bar{\mu} = 4.3$, $t = 4.9934$, $df = 9$ and $p = 0.00035$ (which is less than $\alpha = 0.05$). Hence, this result is **extremely statistically significant**. Hence we reject H_0 .

Conclusion - Section level annotations are more helpful than document level annotations.

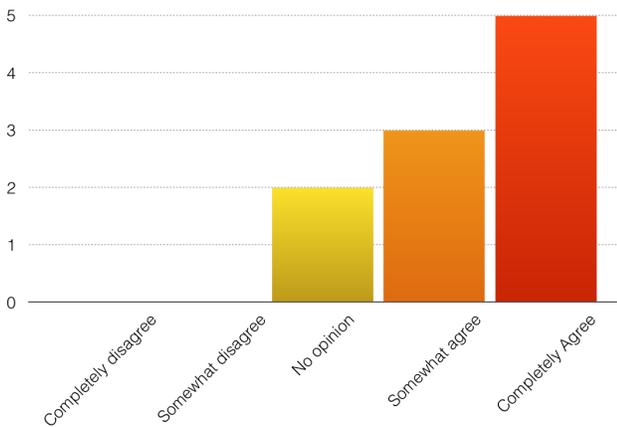


Figure 10. Is the section view more useful than the document view?

Willingness to Share Annotations

We have shown that annotations are effective (especially the section level annotations), but are researchers willing to share annotations? We suspected that researchers might respond differently depending on who they share their annotations with. Hence, we asked them two separate questions. First, “Are you willing to share annotations publicly?”. The left graph

in Figure 11 shows the distribution of responses. We use a one-tail one-sample t-test to verify significance.

H_0 : Researchers are neutral towards sharing their annotations publicly. ($\bar{\mu} = 3$)

H_1 : **Researchers are willing to share their annotations publicly** ($\bar{\mu} > 3$)

We obtained $\bar{\mu} = 3.5$, $t = 1.2457$, $df = 9$ and $p = 0.1222$ (which is more than $\alpha = 0.05$). Hence, this result is **not significant**. Hence we cannot reject H_0 .

Next, we asked them a milder version of this question, “Are you willing to share annotations within your research group?”. The right graph in Figure 11 shows the distribution of responses. We use a one-tail one-sample t-test to verify significance.

H_0 : Researchers are neutral towards sharing their annotations in a research group. ($\bar{\mu} = 3$)

H_1 : **Researchers are willing to share their annotations in a research group.** ($\bar{\mu} > 3$)

We obtained $\bar{\mu} = 4.2$, $t = 3.3425$, $df = 9$ and $p = 0.0043$ (which is less than $\alpha = 0.05$). Hence, this result is **extremely statistically significant**. Hence we reject H_0 .

Finally, we compared the two distributions using a one-tailed two-sample test.

H_0 : Researchers find no difference in sharing their annotations publicly or in a research group.

H_1 : **Researchers are more comfortable sharing their annotations within a research group instead of publicly.**

We obtained $t = 1.5302$, $df = 18$ and $p = 0.0717$ (which is slightly more than $\alpha = 0.05$ but less than $\alpha = 0.1$). Hence, this result is **slightly insignificant**. Hence we cannot reject H_0 with a confidence level of 95%, but we can reject H_0 with a confidence level of 90%.

Conclusion - Researchers are willing to share their responses in a research group. Researchers are probably more willing to share results within their research group rather than publicly (with 90% confidence level).

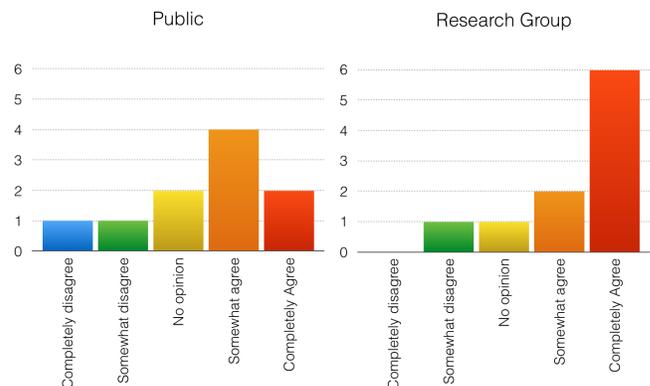


Figure 11. Are researchers willing to share their annotations?

DISCUSSION AND LIMITATIONS

We build Research Exchange and carried out a few experiments in naturalistic and controlled settings to evaluate how effective this tool is in practice. Research Exchange shows promising results for both of our objectives.

Our first experiment tested the effectiveness of collaborative annotations and showed that annotated PDFs are a good solution pedagogically [18]. Collaboratively annotated PDFs also allow discussion of documents which can be a valuable tool for peer review.

Our next experiment showed that the section level view is more effective than the document level view. We decided to design a section-level view since many young researchers complained that they were often stuck on particular parts of the PDF and we believed that fine-grained context-specific annotations will be a helpful resource. Once section-level annotations have been made at a large scale, this might be a good way to identify the most difficult / most important / most controversial parts of a paper. Large scale section level annotations are also likely to paraphrase the original text, potentially solving the trouble while finding papers online. Finally, “Supplementary Material” section level annotations can help identify related work, thereby leading to better clustering of similar documents.

While Research Exchange might be very successful at scale, will people be willing to contribute to it? Our third experiment showed that people are possibly unwilling to share their annotations publicly. This is a big concern and the most likely point of failure. A strong incentive-driven model has to be developed (such as recognition in the community, job prospects, rating systems) to encourage people to spend time and effort writing annotations or sharing their annotations online. Our experiment also shows that people are more willing to share annotations within a research group. This is probably because people personally know the people they are sharing annotations with, and not sharing annotations might lead to friction within the group. This is also why we believe Research Exchange will work in peer-review settings, since people will be obliged to share their thoughts about particular parts of the paper.

Limitations

Research Exchange has several important limitations which need to be addressed before it can be deployed in the real world. We discuss them in this section.

- Adding new papers to Research Exchange is a slow and manual process. An automatic “paper-to-text” solution has to be developed to build a more usable interface.
- Our system lacks several important features like private annotations, support for replies and discussion, deletion of annotations, user accounts etc. These missing features were pointed out during our usability studies and need to be addressed before the tool can be deployed.
- The user interface is not intuitive enough and there lacks the instructions on how it could be used. Therefore, researchers made some mistakes, such as categorizing the annotations not in the intended way. This interferes with the results in some minor ways.
- The sample size is limited which makes the results not significant enough to give more general conclusions.

- Research communities are relatively small. Annotations for most papers could be very sparse. The effectiveness of this project might be limited, unless strong incentive models are built around Research Exchange.

Since what we have for now is still a prototype, there are many user interface improvements possible. There are also many possible functions to be added. We will discuss any future work in the next section.

CONCLUSIONS AND FUTURE WORK

This paper introduces a new way to read paper. In the early design process, we observed the reading behavior of new and experienced researchers to make a preliminary investigation. Then we designed a reading system that allows the entire research community to participate in providing ideas, elaborating the papers and publishing annotations. A key design consideration is on how to produce effective annotations to help readers better understand the contents of papers. We designed a flexible experimental process that allowed participants to freely add and use annotations. Through a period of deployment and testing, we have collected a lot of feedback that will help us improve the performance of our system.

Research Exchange succeeded in getting people to express their understanding and evaluation of papers in detail and presenting them in intuitive ways. However, since this project is still in the prototype phase, we only tested our work with a small number of researchers and some graduate students who just start to read papers. It is important to understand how our approach fits into other contexts and broader scales which can involve the entire research community. We have identified several key areas for future work.

To deploy the product on a larger scale, a few important steps need to be completed. Since research papers are usually available in PDF format, to get more papers into our system, it is essential to allow automatic conversion of paper format from PDF to the JSON format in our system, potentially using ParsCit Structured XML. Moreover, the interface needs to be modified to address the limitations - primarily support for accounts; private annotations using Linux model of user, group, public; and support for replies and discussion. To facilitate this development as well as to improve the public visibility of this tool, we hope to collaborate with the OpenReview team at UMass Amherst over the next few months. Once this tool has been adopted by several people and contains several thousand annotations, it can be an excellent test-bed to carry out data analysis on people’s paper reading patterns, building better software to cluster similar papers together and improve search for technical information.

REFERENCES

1. Waleed Ammar, Dirk Groeneveld, Chandra Bhagavatula, Iz Beltagy, Miles Crawford, Doug Downey, Jason Dunkelberger, Ahmed Elgohary, Sergey Feldman, Vu Ha, and others. 2018. Construction of the Literature Graph in Semantic Scholar. *NAACL HLT 2018* (2018), 84–91.
2. Anjo Anjewierden, Hannie Gijlers, Bas Kolloffel, Nadira Saab, and Robert De Hoog. 2011. Examining the relation between domain-related communication and

- collaborative inquiry learning. *Computers & Education* 57, 2 (2011), 1741–1748.
3. Christine Betts, Joanna Power, and Waleed Ammar. 2019. GrapAL: Querying Semantic Scholar's Literature Graph. *arXiv preprint arXiv:1902.05170* (2019).
 4. Chandra Bhagavatula, Sergey Feldman, Russell Power, and Waleed Ammar. 2018. Content-Based Citation Recommendation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Vol. 1. 238–251.
 5. Nian-Shing Chen, Daniel Chia-En Teng, Cheng-Han Lee, and others. 2011. Augmenting paper-based reading activity with direct access to digital materials and scaffolded questioning. *Computers & Education* 57, 2 (2011), 1705–1715.
 6. Michael D Ekstrand, Praveen Kannan, James A Stemper, John T Butler, Joseph A Konstan, and John T Riedl. 2010. Automatically building research reading lists. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 159–166.
 7. Katharine E Hubbard and Sonja D Dunbar. 2017. Perceptions of scientific research literature and strategies for reading papers depend on academic career stage. *PloS one* 12, 12 (2017), e0189753.
 8. Ziming Liu. 2005. Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of documentation* 61, 6 (2005), 700–712.
 9. Ehsan Mohammadi, Mike Thelwall, Stefanie Haustein, and Vincent Larivière. 2015. Who reads research articles? An altmetrics analysis of M endeley user categories. *Journal of the Association for Information Science and Technology* 66, 9 (2015), 1832–1846.
 10. Les Nelson, Christoph Held, Peter Pirolli, Lichan Hong, Diane Schiano, and Ed H Chi. 2009. With a little help from my friends: examining the impact of social annotations in sensemaking tasks. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1795–1798.
 11. Kenton O'hara and Abigail Sellen. 1997. A comparison of reading paper and on-line documents. In *CHI*, Vol. 97. 335–342.
 12. Juan Ramos and others. 2003. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, Vol. 242. 133–142.
 13. Wendy Roldan and Jennifer A Turns. Graduate Students Learn to Read Academic Research. (????).
 14. Noah Siegel, Nicholas Lourie, Russell Power, and Waleed Ammar. 2018. Extracting scientific figures with distantly supervised neural networks. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. ACM, 223–232.
 15. Gemma Walsh. 2016. Screen and paper reading research—A literature review. *Australian Academic & Research Libraries* 47, 3 (2016), 160–173.
 16. Noreen M Webb. 1989. Peer interaction and learning in small groups. *International journal of Educational research* 13, 1 (1989), 21–39.
 17. Amy X. Zhang, Michele Igo, Marc Facciotti, and David Karger. 2017. Using Student Annotated Hashtags and Emojis to Collect Nuanced Affective States. In *Proceedings of the Fourth Annual ACM Conference on Learning at Scale: Poster Publication (L@S '17)*. ACM.
 18. Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. 2012. Successful classroom deployment of a social document annotation system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1883–1892.