Let's Read: Online paper-reading platform with crowd-generated semantic highlights

Junsoo Park*, Sang-Gyun An*, Youngbo Shim* (*: co-first author)
KAIST
{parkjunsoo91, sang-gyun.an, aram6207}@kaist.ac.kr

ABSTRACT
We introduce Let's Read, an online paper-reading platform where a group of novice researchers read a paper together with helpful visual cues naturally generated by themselves. It tries to solve problems of novice researchers when reading an academic paper, which are a poor time-efficiency and a lack of critical view. By providing a multi-color highlighter to express different semantics such as 'normal highlight', 'like', 'dislike', and 'don't know', and shared highlights generated by a group, it helps novice researchers to make and share helpful visual cues on a paper in real time. An in-class user study shows that it helps with writing a critical reading assignment of papers.

INTRODUCTION
Reading an academic paper efficiently is crucial for researchers to be accustomed to a discipline and develop his or her research topic. Thus, novice researchers, such as first-year graduate students, are commonly required to read papers as many as possible. Thanks to Internet, a massive amount of papers is readily available via, for example, ACM Digital Library [1] and Google Scholar [2]. However, when reading a paper, novice researchers usually experience a poor time-efficiency and difficulty in having a critical view. It is because they are unfamiliar with a discipline and have little background knowledge.

To deal with this problem, we have found that some novice researchers prefer reading a paper with annotations made by expert researchers, such as professors or senior graduate students. At a glance, they can figure out which part is important and can see which part the expert researcher likes or dislikes, which could help to build a critical view of a paper. However, there are obvious limitations. At first, an expert is rare and expensive. Moreover, it is hard to guarantee their motivation to help novice researchers. Also, a hard-copy paper lacks scalability.

We introduce Let's Read, which offers a semantic highlighter and group-based shared highlights. The highlighter uses four colors and each corresponds to different semantics such as 'normal highlight', 'like', 'dislike', and 'don't know'. Also, as for displaying shared highlights, we implemented an aggregation algorithm. We designed and implemented the system based on feedbacks from a pilot study, and deployed it online around for 3 days. We had 13 users including three of us and found potential and limitation of our system at the same time. Lastly, based on findings from the deployment, we propose possible improvements.

BACKGROUND AND RELATED WORK
There are few services where users can read a paper together such as Mendeley [7], EndNote [8], Google Drive etc. However, there is no aggregation algorithm which is critical for a scalability issue of crowd. Also, Kindle [9] is an app providing social highlights, but it doesn't use colors to differentiate semantics of each highlight.

SYSTEM
Following paragraphs include a workflow, design implications, an interface, an aggregation algorithm, and implementation of Let's Read.

Workflow
The system uses a straightforward workflow (Figure 1), which consists of several steps: 1) uploading a paper and creating a group, or joining an existing group, 2) making highlights with a multi-color highlighter, 3) voting on others' highlights, and 4) aggregating all highlights. In fact, step 2 and step 3 can be executed simultaneously. Rather than introducing another separate task for voting, we tried to use a natural behavior of user, which is highlighting over others' highlights. Thus, when user makes a highlight over an existing one, the system counts it as a vote (Figure 1.a). Every Time each user makes highlights, computer algorithmically aggregates all highlights (Figure 1.b).

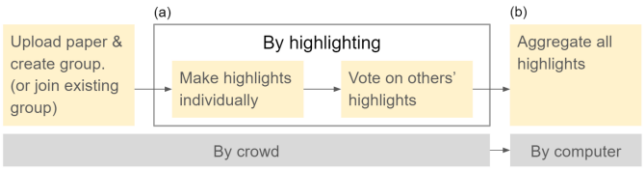

Figure 1. A workflow of Let's Read.

Design Implications
We ran a pilot study using our low-fi system prototype (Figure 2). The system was merely a graphic mockup, of which features are fake or just existing features of a commercial pdf reader (i.e., Adobe Acrobat Reader). With the low-fi prototype, three participants (P1, P2, P3) were given a first page of a HCI paper and required to read and understand it. Participants were instructed to use colors to highlight a content if they need to make more semantic highlights. For this, we prototyped a three-color highlighter, of which color corresponds to 'normal highlight', 'like' and 'dislike'. It is based on a guideline for casual critical review used in CS492 in KAIST [3]. Also, participants were able to see others' highlights, which we had prepared in advance by ourselves.

By requiring participants to use thinking-aloud technique [6], we could get some valuable qualitative feedbacks on the low-fi prototype. At first, all three participants said that they felt more confident in reading with others' highlights. However, surprisingly, they tended to be irritated easily when they saw other's highlights which, he or she thinks, is irrelevant. For example, P2 said "I never understand why people made a highlight here!". Also, from a post-hoc interviews, we found that participants were in favor of the concept of using multi-color highlighter for semantic highlights. However, they were reluctant to use it actively because of a critical interface usability issue. P3 said "distance of mouse travel hinders me from changing color of highlighter". Lastly, P1 and P2 said that they needed another color to flag a content which they 'don't know'.

Based on pilot study, we drew three design implications:

- I1: Aggregation should filter out irrelevant highlights
- I2: Long distance of mouse travel hinders users from using multi-color highlighter
- I3: Users need to express 'don't know'

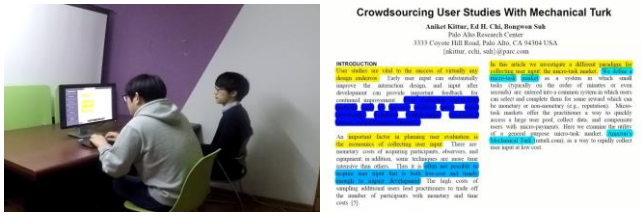

Figure 2. Pilot study uses low-fi prototype of the system. Each participant was given a HCI paper to read with our system (left) and instructed to use highlighter with colors to make highlights more semantic (right).

Interface

The system interface mainly comprises three parts: a login field, a paper view, and a hovering color palette (Figure 3). To use our system, users are required to log in with university email account. It is deliberately designed for restraining anonymous sabotage. After logging in, users can see others' highlights on paper. As to the multi-color highlighter, according to the design implication I3, the system provided four colors which corresponds to 'normal highlight', 'like', 'dislike', and 'don't know' (Figure 3.2.). Based on the design implication I2, a color palette follows scroll bar to shorten mouse travel for changing color. Also, users can change color by using a shortcut key which is assigned for each color (Figure 3.2.). Finally, by using dragging mouse action, users can make and share semantic highlights on the paper.

Aggregation Algorithm

In regard to quality control, according to the design implication I1, we revised our aggregation algorithm to show more agreeable highlights generated within a group, and hide others which are irrelevant. In order to do that, the system counts number of votes for highlights of each word in paper, and hide them if their number of votes are below a threshold (Figure 4). To keep the threshold reasonable with scalability of the system, the threshold is dynamically adjusted according to the largest number of votes of a word in a paper (Figure 5).

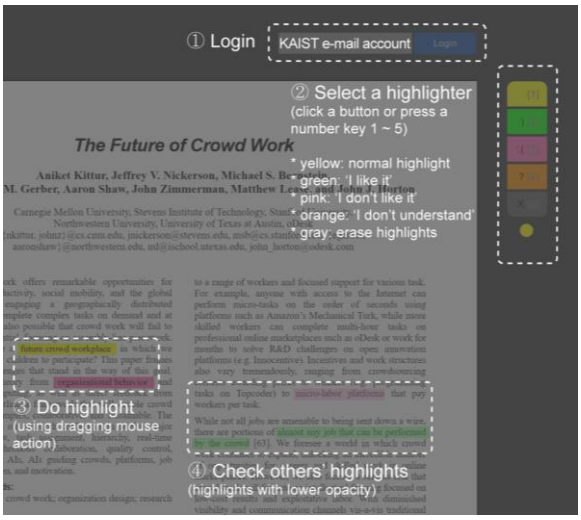

Figure 3. The overview of system interface, where user can login using his or her university email account (1), select a color of highlighter (2), make highlight on a paper (3), and see others' highlights.

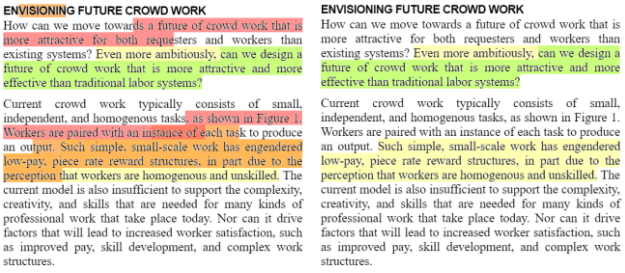

Figure 4. Left: aggregated highlights without a threshold. Right: aggregated highlights with a threshold.

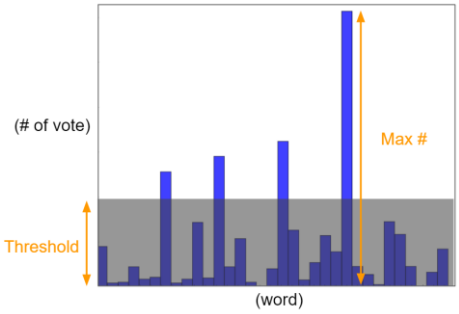

Figure 5. The threshold for aggregating highlights, which are generated within a group, is proportional to the maximum number of highlights of a word.

Implementation

To implement Let's Read, we use HTML, CSS, Javascript, and Flask (a python microframework for the server-side). We tried to keep the system as simple and swift as possible, thus we decided not to use other library and API except for JQuery and Flask for implementation. In terms of view, to

update and show individual highlights and others' highlights respectively, the system used separate layers for them by using basic HTML and CSS.

EVALUATION

User study

To evaluate the effectiveness of the semantic highlights and highlight sharing features of Let's Read, a user study was conducted in a class setting. The class periodically gave an assignment to read a specific paper and write a reading response. The reading response requires students to write 3 paragraphs: first, summarize the main idea, second, write things that he learned, and third, write any methodological, technical, logical issues and possible improvement of the work [3]. These tasks require a good understanding of the paper and the ability to critically assess the details of the paper, which are the activities that Let's Read aims to support. The platform was open to access for the assignment's due day. We hypothesize that the students who use the platform perform better on the assignment in terms of depth of understanding and quality of analysis.

Result

The user study was conducted for a 17-page long paper, "The Future of Crowd Work". Advertisements were made in the class website, but no person was directly recruited. At the end of the day, a total of 3 people used Let's Read and used it to write their reading responses, contributing a total of 99 highlights. The highlights constitute of 64 'normal highlight', 21 'likes', 7 'dislikes', and 7 'don't knows'. The observation that people used normal highlights more extensively indicates that highlighting the main points and keywords of passages was the most common activity for paper readers. People were more selective when they were pointing out the 'dislikes' which requires careful, critical thinking. However, the data set is too small to draw any statistically significant conclusion.

Qualitative survey

To gain a further understanding of the effects of the highlights on the reading responses, an online survey asking the qualitative results was conducted. The 3 people who used the platform were surveyed and were asked about the time it took to read the paper and complete the assignment, about the effects of multi-color semantic highlights and about the effects of presenting the crowd's highlights.

Question	Answer(avg.)
How long did it take you to read a paper? (with our system)	103 min.
How long does it take you to read a paper on average? (without our system)	110 min.
How long did it take you to write a reading response? (with our system)	42 min.
How long does it take you to write a reading response on average? (without our system)	27 min.

Table 1. Survey results about the time for reading and writing a reading response.

Table 1 shows that the system did not reduce but slightly increased the paper reading time, which could be due to the overhead of highlighting activity. However, the reading response writing time decreased significantly. This hints that the system helps reduce time it takes to review the good and bad points that the user has pointed out while reading.

Question	Answer(avg.)
How helpful were multi-color highlights when reading the paper?	3.0/5
How helpful were multi-color highlights when writing a reading response?	4.7/5

Table 2. Survey results on the effectiveness of multi-color highlighting.

Results in Table 2 shows that the multi-color highlights were only marginally effective in reading and understanding the contents of the paper, but was significantly more effective in writing a reading response. The comment "I can easily recall the parts which I liked or disliked", suggests the highlight's role of a read-later note. On the other side, comments that the highlighting activity "disturbed the natural reading flow" and sometimes "ambiguous points made it hard to decide which color to use" shows the downsides of the highlight options.

Question	Answer(avg.)
How helpful were others' highlights when reading the paper?	4.0/5
How helpful were others' highlights when writing a reading response?	3.3/5
Was the amount of others' highlights enough?	3.7/5

Table 3. Survey results on the effectiveness of presenting the crowd's aggregated highlights.

Mixed responses were shown about the presentation of crowd made highlights. As shown in Table 3, other people's highlights were not much appreciated when writing a reading response. While some people thought positively, like "trying to understand other people's highlights made me think about the issue deeply that I would have otherwise missed", some people said "I tended to follow other people's highlights unconsciously," suggesting hindrance to autonomous reading. This suggests the need to provide a more concrete reason to trust or not a given highlight made by others.

DISCUSSION

Insights of user study failure

For an effective evaluation, highlights from start to end page of the paper is needed. Users must read the paper wholly to provide a valid data set. Also, these readers should be motivated because participants that are forced to read the paper won't pay attention on the paper and may highlight on meaningless parts.

We figured out that all instant users didn't do much highlights and even done highlights were meaningless to the context, which are thought that they have just tried the highlight function for fun.

On the other hand, we could think that not enough users from the assignment channel may due to tight deployment date. The system was noticed to the assignment channel just one day before the due date, and thus potential users might hesitate to use unfamiliar tool in a hurry.

Analysis on general findings

From the qualitative survey and discussion, some potential issues and goods of the system were revealed. All these findings look quite general that other common users might agree on same feelings while using the service, so we share these insights.

Consuming time for reading

We have observed that the average paper reading time of users have slightly increased (Table 1). Also, the assignment working time was reduced. So even actual reading time have increased, we could think it as a slight tradeoff with a focused reading on a paper. Also, though the reduced assignment doing time was mostly benefit from the multi-color highlight function itself, well understanding of the paper might have contributing a part of doing the assignment. Based on this results, we could expect Let's Read will indeed support to help novice researchers by providing them thorough reading on paper with efficient time use when it is deployed in public.

Optimal highlight density

One trend of highlight was that the density varies by the context of paper. For instance, abstract and introduction part of the paper were significantly highlighted much with various colors (Figure 6). This is a natural behavior because usually scientific paper's main idea is concentrated in the abstract and introduction part. The problem is that user's might need more group highlights in the sparsely highlighted parts of the paper.

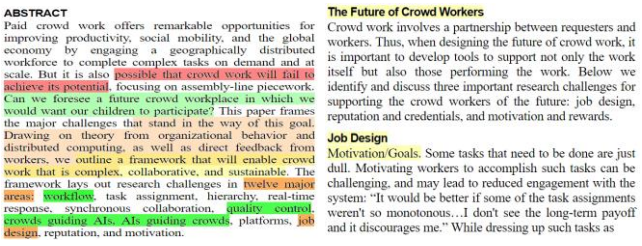

Figure 6. Density variation between abstract part(left) and discussion part(right)

Also, from the qualitative survey, users tended to have different satisfaction from the same group highlights. That is, the optimal highlight display setting varies by individuals. Some users might prefer sparsely highlighted ones to concentrate on the passage itself and only find out highly weighted keywords from the group highlight, while others might want to use group highlights as a summary of a paragraph using dense highlights. Furthermore, users might have a dynamic preference regarding on the context of the paper. Since Let's Read has a constant threshold to aggregate the highlights, it is hard to serve all user tastes.

Limitation

There is an intrinsic problem coming from the design itself. Early users of the system couldn't benefit from the group highlights. Since the highlights are accumulated by the users over time, first user will start highlighting on a yet untouched paper. This not only take away the chance of user to trace other's work but even discourage them because of unfairness. Eventually it might end up with nobody trying to start reading. This problem could be resolved by providing other functionalities attractive to the users, so they will to use the system even without the group highlights. On the other hand, compensating the user by honoring he/she on a leaderboard that is displaying early contributed readers might draw users' attention.

Possible improvements

From the analysis on the system, we suggest two functions to improve the system.

Highlight slider

From the observation that user preference varies by individuals, we concluded that fixed visualization might not work. As a solution, we present the idea of highlight density controller called highlight slider. Since Let's Read aggregates the highlights from group members and filter out less overlapping highlights based on the threshold constant, the display of group highlights could be visualized dynamically if the threshold is modifiable. By providing the controller as a slider GUI, user might see the passage by their own preference in real time. The design of this idea is depicted in Figure 7.

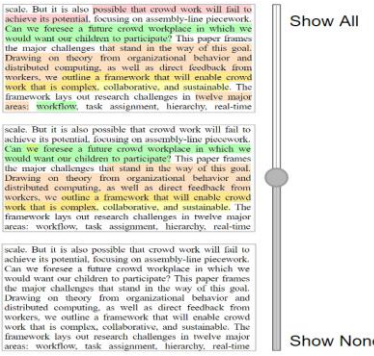

Figure 7. Highlight slider concept

Reading progress bar

Reading a paper fully is a hard work, and this barrier had discouraged many instant users reading the paper through our system. There have been already lot of trials to motivate the users using a game component, which is called gamification [5]. Progress bar is a frequently used component to feedback and visualize user's goal. Inspired by this, we introduce reading progress bar, which shows how much paragraph users have read based on paragraph's users have highlighted.

REFERENCES

1. ACM Digital Library. (n.d.). Retrieved December 22, 2016, from http://dl.acm.org/
2. Google Scholar. (n.d.). Retrieved December 22, 2016, from https://scholar.google.co.kr/
3. Kixlab. Reading Response. 2016. Retrieved December 22, 2016, from http://kixlab.org/courses/crowdsourcing/reading-response.html
4. Keshav, S. 2007. How to read a paper. ACM SIGCOMM Computer Communication Review, 37(3), 83-84.
5. Deterding, S., Dixon, D., Khaled, R., & Nacke, L. 2011, September. From game design elements to gamefulness: defining gamification. In Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments (pp. 9-15). ACM.
6. Kixlab. Low-Fi Prototype. 2016. Retrieved December 22, 2016, from http://kixlab.org/courses/crowdsourcing/final-project.html#4
7. Mendeley. (n.d.). Retrieved December 22, 2016, from https://www.mendeley.com/
8. EndNote. (n.d.). Retrieved December 22, 2016, from http://endnote.com/
9. Kindle cloudreader. (n.d.). Retrieved December 22, 2016, from https://read.amazon.com/