Limited Text Count Attention Study

Tobias Harnacke
School of Computing
Clemson University
tharnac@clemson.edu

Adam Boyd
School of Computing
Clemson University
atboyd@clemson.edu

Andrew T Duchowski
School of Computing
Clemson University
andrewd@cs.clemson.edu

ABSTRACT
The purpose of this experiment was to use eye tracking to see when a person decides whether a paragraph of text is relevant to his/her interests or not. We could use this information to improve web usability, especially in regards to limited text social media. The goal of the experiment was to see if there was a significant difference between changing paragraph character count and how far a person would read into a paragraph. We failed to reject the null hypothesis because we were unable to find significance when changing the character count. This could show that people only read a certain amount of text before making assumptions of the remainder. The means of the values were slightly different based on character count, but not enough to be significant.

Author Keywords
Eye tracking, web design, contextual snippets.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Twitter, an online media company, has offered a solution to the problem of 'too much information on the Internet'. Twitter incorporates a policy that limits character count of messages which forces users to get directly to the point they are trying to make. Web systems like Twitter are also used for entertainment purposes, not solely information seeking as you would use a web search engine. Our study focused on limiting character count of paragraphs and having participants seek information within those paragraphs.

Hypothesis
We ran our experiments to see if there was a significant difference between varying paragraph sizes. We assume the null hypothesis that the amount of characters a person reads is not affected by how long a paragraph is when seeking information. The experiment was designed to find significance when changing this factor, character count. We felt that it was possible that the character count read by the participant could be roughly equivalent to Twitter's character limitation of 140 characters. We considered that this could be the average character count required by most individuals to assume the general content of a paragraph.

RELATED WORK
The media company, Twitter, in its advent wanted to adopt the 140 character limit that is incorporated into SMS used by cellular phones. Because of these SMS limitations, Twitter wanted to capitalize on the people who used SMS to communicate by setting a text limit. Though more and more people are using smart phones - most of which have messaging applications to circumvent SMS limitations - Twitter has adhered to the character limit. Intuitively, this may seem like a negative, however, it is shown to be one of the more appealing aspects of this type of social media.

In this paper, we made use of eye tracking techniques to determine when a person assumes to know the general content of a snippet without having read it in its entirety. Given a list of paragraphs, how many characters are read until an article is concluded to contain the correct information asked for in the task? Once the person knows that the information they seek is not likely in the remainder of the paragraph, what do they do? And when does this occur?

There have been many studies regarding web search engines and their efficiency [4, 6]. While we used this knowledge to our benefit, our study expanded on the idea in the context of limited text social media. We still incorporated an information seeking task, but the information required to complete the task may be obscured by many sections of similar content which forces the reader to read into the paragraphs in order to determine if the text could contain the information. While we expected this to produce a heavy left-side distribution of fixations, we assumed it to not have as prominent of a golden triangle or "F" shape as shown in web search engine studies [7]. The reason for this is because participants needed to continue their search until they found the paragraph containing the task's objective.

A study on web search engines, Cutrell, et al. [4] gave insight to what we expected to find in this study. The particular study focused on navigational and informational tasks for a web search. He determined the fixation length on particular objects of the search: the title, the actual snippet, and the URL. Our study did not have all of these other fields but focused solely on a snippet of information. Using Wooding's work in fixation mapping, we expected to elaborate on the focal points that participants have fixated on the screen [8]. Perhaps participants fixate on certain

paragraphs longer than others. The language used in our study may have an effect to cause more attention to some details over others.

Scan-path studies gave us insight into what we can expect from eye-tracking data from participants who have engaged in a reading task [5]. Using Anderson's study on how the human brain is working while the eyes are in motion, especially in regards to visual search, we hoped to find some meaningful results [1].

Recording of participants' gender was important based on previous study findings. Lorigo, et al. [6] found that scan-paths of males and females differed in the order in which they scanned a document and males would look at more results than females would.

METHODS

Apparatus

The tasks' text was built into an image which resembled a web page and displayed in Tobii Studio. Eye tracking was done using the Tobii 1750, which is a 17" monitor with cameras embedded, and the data was recorded using the Tobii Studio suite mentioned above. This particular eye tracker, the Tobii 1750, samples the intended position of the participants' eyes at 50Hz and has an accuracy of 1°. The Tobii 1750 was calibrated to each participant using a 5-point calibration grid prior to the task.

Participants

We had 12 participants, 11 males and one female, in our study. All participants had at least an eighth grade reading comprehension level of English, however four of the participants were from Germany and therefore not native English speakers. Two of the participants had a sampling rate lower than 70% so they were excluded from the final results. The task took roughly five minutes to complete after which we offered to show participants their data if they wanted to see it and we explained the purpose of the study for those interested.

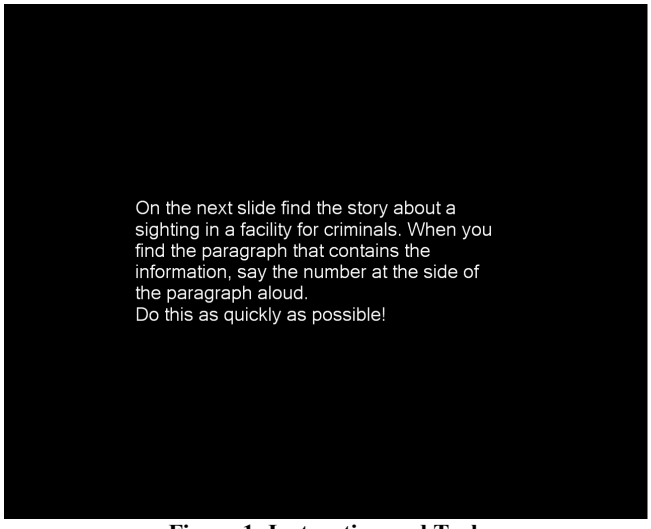

Figure 1: Instruction and Task

Experimental Design

We designed the study to help understand the ways a person would seek information when many similar topics are all presented on a single page in list form. Each page contained a list of paragraphs with six pages in total.

The study had one task for each of the three lists of paragraphs. Each list had paragraphs of only length 200, 350, and 500 characters. We had two of each of these lengths for a total of six full-page length lists. The number of paragraphs varied for the number of characters per paragraph. For the list of 200 character length paragraphs, there were six total paragraphs for participants to read, five total paragraphs for the 350 character count list and three total paragraphs for the 500 character count list. Participants had to read to complete the information seeking task and once it was completed, they reported the number aloud. We told participants to report the number in order for users to read accurately. The data regarding the accuracy of answers are not important for this study, but it was important for participants to focus on accuracy.

Paragraphs were separated from each other in order to accommodate specific regions of interest used when extracting data. Only one paragraph contained the information needed to complete the experiment and it was located somewhere between the middle to the bottom of the list in an attempt to draw in more eye-tracking data. The paragraph for each list was fixed for all conditions.

Participants were urged to complete the task as quickly as possible, but not at the cost of being inaccurate in their answer. To make sure of this, we told participants that we were recording time for each task. The data of time were not relevant to our study.

1. My daughter was sleeping in my bed with me one night. She has always seen things. That night I had my closet light on. The light reflects directly across to my TV. She said she looked into the TV and there was a little girl standing in my closet door. She tried to wake me up but I was knocked out. I remember feeling her shaking next to me but I just couldn't wake up.

2. In Gaffney, I was sitting in my mom's friend's house and me and my friend were sitting on the bed. We hear heavy boot-like stomps in her hallway. We get up and look in the hallway. We could stare in the hallway and see nothing but hear it coming closer and closer until it stopped right in front of us. We shot straight down the hall and outside.

3. I had a friend that lived in Blacksburg. He said that they had not lived there very long and he opened the door to go upstairs and when he looked up the stairway there was a man at the top of the stairs with wide shoulders and no face blocking him from coming up the stairs. Him, his son, and wife all slept in the living room that night and quickly left the next day.

4. I took a ghost tour of the city one day around dusk a couple of years ago. The tour included the old city jail, which was just during the Civil War to house federal soliders. Upon arriving at the jail I saw a fiqure of a man in the second story window and turned to my friend that we were in a surprise scaring.

5. A glowing figure of a young girl was spotted watching a friend of mine sleeping on the couch in his apartment around Colonial lake. When he awoke the ghost walked through the wall into his bedroom and then vanished. We found out that this building used to be a hospital a long time ago.

Figure 2: Stimulus - List of paragraphs with 350 characters

Instead of discarding saccade data and focusing solely on fixation data, we needed to watch each of the regions of interest (in this case, each textual paragraph clearly separated to be accurate within at least 1°) and determined how long and far the participant read each region. Our study expected saccades to occur when a participant seeks information to complete their task.

The experiment was within subjects so that each participant was exposed to 200, 350, and 500 character length paragraphs. In order to eliminate training effect we utilized the Latin Square Method and varied the order in which participants viewed the respective character length paragraphs.

In order to minimize scanning for keywords as a mean to find information, we made sure to not use the exact same wording in the task description as it would be found in the paragraphs.

Procedure

The experiment was carried out in a single session. Participants arrived in the eye-tracking lab and were read instructions that told the participant the purpose of the experiment. The session began with a 5-point calibration of the eye-tracker to the participant. Once the Tobii 1750 was calibrated, the user was given instructions to find specific information from a list of paragraphs of approximate length of 200, 350, and 500 characters. Each page contained a list of paragraphs of stories about ghost sightings. After completion of the tasks, participants were then presented with an instruction for the next task. We had two tasks and lists of each character counts so that there were six tasks of information seeking. For each task, we recorded the time participants took to complete it.

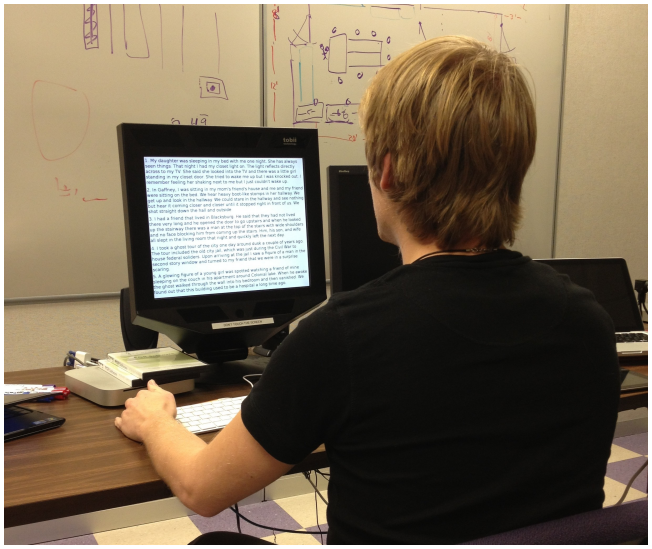

Figure 3: Participant and Tobii 1750 Eye Tracker

In order to confirm they accomplished their task successfully, we asked the participants to verbally announce the number of the paragraph in which they found the correct information. Once completed, we asked the participant if they have any additional questions and once the questions were met with sufficient answers they were thanked and dismissed.

Once all experiments had been carried out we aggregated the data based on gender because of the results of a previous study showing that males and females have differing scan-paths when seeking information [6]. For this particular case we found no difference between the scan-paths of the genders. This is likely because of our sample size of a single female.

RESULTS

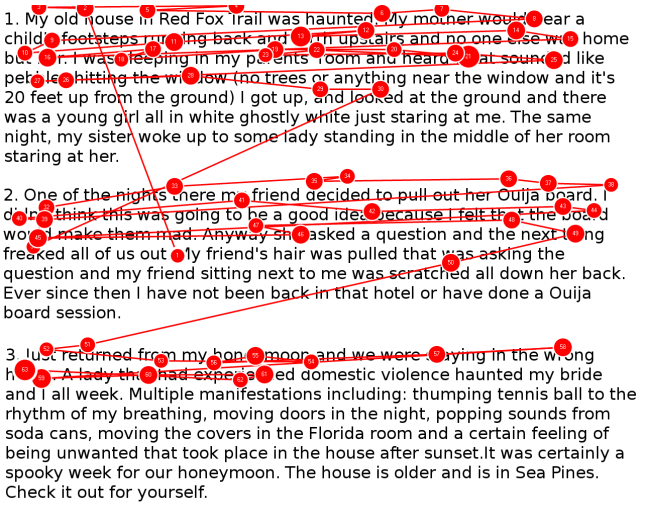

Figure 5: Gaze plot of a participant's reading

As we analyzed the data, an obstacle was finding the best way to figure out how many characters a participant had read. We found that manually reviewing the scan-paths recorded by Tobii Studio was the best method to achieve that. Just using automated scripting would have resulted in inaccurate character count information because of irrelevant eye movements and limitations to the eye tracker (blinking, calibration issues etc.). We recorded the amount of characters read by each participant for each paragraph. We ignored the paragraphs which contained the information needed in order to complete the task, because the position of the information in the paragraph would have affected the amount of characters read. Furthermore, we ignored data for paragraphs not read and ones where it could not be determined how many characters were read due to participants scanning the paragraph vertically for key words.

We also used the Tobii studio statistics tool to gather fixation data for specific areas of interests (AOIs). The

paragraph where users said was the one which contained the data was selected as the AOI. We retrieved the numbers of within the AOI as well as outside the AOI to find the fixation duration for both. As the heat map in figure 6 emphasizes, participants would often revisit or spend more time fixating on the paragraph which contained the information needed to complete the task. This may imply a process in the human mind critically analyzing the material that was just read. This related to Anderson's study of higher level cognition in visual search tasks and visual attention [1]. The sums of the means of the duration of fixations of all but two cases were greater for the paragraphs that contained the data. That means that all other paragraphs in a list combined were fixated on less than the single paragraph containing the data. Since studies show a correlation between fixation time and thought process, it is safe to say the most thought provoking aspect was the area that corresponded with the task [1]. This was not the purpose of our study, but it was interesting to find more data to substantiate others' work in eye tracking.

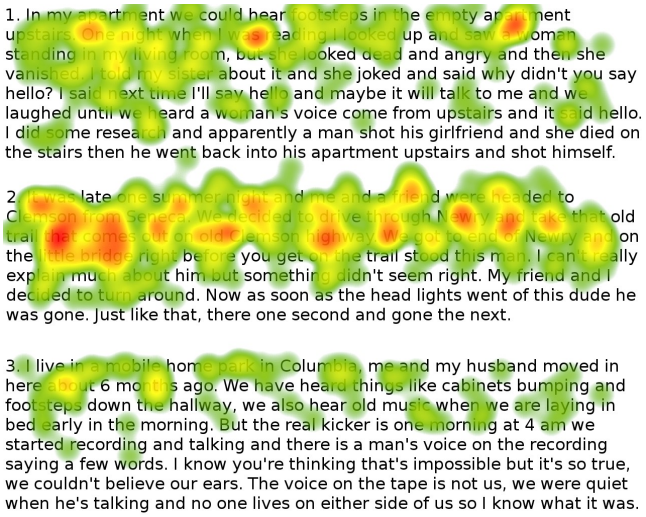

Figure 6: Heat map of stimulus

For our primary study, a single factor (paragraph's character count) ANOVA of characters read was used to evaluate the significance of our data. Our ANOVA results show that it was not significant $F(2,27) = 1.85$, $p > 0.05$ (actual value: $p = 0.18$, n.s.). This information is displayed in figures 4 and 7.

For the 200 character count paragraphs we had a mean of 162 characters read with a standard error of 10%, for the 350 character count paragraphs we had an average of 186 characters read with a standard error of 19.3%, and for the 500 character count paragraphs there was a mean of 206 characters read with a standard error of 19.4%. This information is shown in figure 8. We failed to reject the null hypothesis, due to these results.

	200 characters	350 characters	500 characters
Standard error	10.012150471334	19.2672179567133	19.4081633796166
Average characters read	161.28210978836	185.082222222222	206.275462962963

Figure 8: Table of means and standard error

CONCLUSION

The results showed that the null hypothesis could not be rejected. This means that participants would not read an entire paragraph before looking for the information in another paragraph but rather skip ahead when they assumed that the paragraph they were currently reading was not containing the desired information.

Limitation on character counts in the context of social media therefore is likely a good way too minimize the amount of text an individual has to read through. Increasing the maximum character count would not have a positive impact in means of helping people find pertinent information.

In terms of general web development, these results suggest that the first couple of hundred of characters are important to portraying the entirety of the snippet's information. This would allow people to locate information faster and easier in an online environment.

Limitations

We understood that several aspects of our study contained certain limitations. In the design of our study, we had a varying number of paragraphs for each list of differing sizes of character counts. The 200 character count list had six, the 350 character count list had five, and the 500 character count list had three. We did this to make sure the screen was filled with text. The problem was that we mistakingly brought this in as an additional independent variable when calculating the time to completion. We were forced to decide whether to keep the total character count roughly equivalent or to keep the number of paragraphs equal.

We found that some subjects, rather than reading a paragraph, would just scan the text randomly for key words to find the information. Upon analyzing that data, we discarded the results for those particular cases because we were not able to determine the characters read.

Some of our data could have been skewed by the fact that some participants forgot the task objective and therefore read through all of the paragraphs. If this did not occur, this would alter the data to an even less significant effect of changing paragraph lengths. We decided to leave these results in instead of marking them as outliers.

While our study focused on a task to locate information, Twitter is not used for this purpose. Twitter is most commonly used to get updates on people or events that a person is already interested in or to leisurely forage for information based on an interest shared by producers and consumers of this information. Therefore, in a real-life scenario, Twitter would not be used for information foraging as our study did.

Future Work

Due to our limitations, an additional study could be performed keeping the number of paragraphs in all lists constant. This would allow for time to completion to be analyzed as part of the study as well. A study using greater differences between paragraphs could also lead to significant results. Perhaps a program could be created in order to reliably determine the exact character count read, because our manual method may have been less accurate and has been more time consuming.

ACKNOWLEDGMENTS

We thank all of the volunteers who participated in this study and completed our tasks on their own time with no compensation other than our thanks. We would also like to thank Blair Shannon and Beckie Nelson for helping us learn the methods of statistical analysis used in this experiment.

REFERENCES

1. Anderson, J. R., Matessa, M., Lebiere, C. ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction, 12*, 439-462.
2. Broder, A. A taxonomy of web search. *SIGIR Forum 36*, 2(2002), 3-10.
3. Buscher, G., Cutrell, E., Morris, M.R. What Do You See When You're Surfing? Using Eye Tracking to Predict Salient Regions of Web Pages. In *Proc. CHI 2009*, 2009, 21-30.
4. Cutrell, E., Guan, Z. What Are You Looking For? An Eye-tracking Study of Information Usage in Web Search. In *Proc. CHI 2007*, 2007, 407-416.
5. Josephson, S., Holmes, M.E. Visual attention to repeated Internet images: Testing the scanpath theory on the World Wide Web. *http://www.factone.com/figures/VisualAttentiontoRepeatedInternetImages_final.pdf*
6. Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., and Gay, G. The influence of task and gender on search and evaluation behavior using Google. *Info. Processiing and Management: an Int'l Journal.* 42, 4 (2006), 1123-1131.
7. Qvarfordt, P., Biehl, J.T., Golovchinsky, G., Dunningham, T. Understanding the Benefits of Gaze Enhanced Visual Search. In *Proc 2010*.
8. Wooding, D.S. Fixation Maps: Quantifying Eye-movement Traces. In *Proc. ETRA 2002*, 2002, 21-33.