Getting the Gist: A Model for Presenting Text at Varying Levels of Granularity and its Effect on Reading Comprehension

Jare Fagbemi
Stanford University
Stanford, CA
ofagbemi@stanford.edu

Haley Kong
Stanford University
Stanford, CA
hgkong@stanford.edu

Peter Doyle
Stanford University
Stanford, CA
pwbdoyle@stanford.edu

ABSTRACT

We designed, implemented, and tested a system, entitled Gist, for delivering articles in a format geared toward giving readers a general sense, or gist, of an article's content. Articles were partitioned into several sections by their authors. These authors then drafted short and medium length versions of these sections. A reading interface displays each of these sections sequentially in their short form and provides a mechanism that allows users to toggle each individual section between its short, medium, and long (unedited) form. We tested articles read in this interface against articles read in a standard online news reading format using a within-subjects comprehension test with twenty-four individuals. Comprehension tests consisted of three overview, three focus, and five detail questions. We found that users of the test interface performed better on overview tasks, especially when reading non-news articles, performed worse on focus tasks, and performed about the same on detail tasks.

Author Keywords
Online news, reading comprehension, text granularity

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces. - Graphical user interfaces

INTRODUCTION
As it stands, it can be difficult for readers of online content to get much out of an article under short time constraints [5]. These restrictions, along with upper bounds on the level of commitment readers have to certain news sources or to an end-goal of "staying informed" makes it all too easy for readers to miss critical or otherwise valuable information simply because they don't have the time to read full-length articles.

We decided to approach a solution to this problem by presenting readers with multiple versions of the same content, each version being of varying length and detail.

Techniques for automated text shortening have been in development for more than half a century [3]. Past work has experimented with shortening articles via machine-learning powered algorithms—one project, Summly, an application that summarizes entire news articles, was acquired by Yahoo for a reported $30 million in 2013.

Several APIs for summarizing large blocks of text are freely available online. We found, however, that the results yielded by these services were often generative—existing text couldn't be refactored into entirely new, more concise text.

For this reason, we asked authors at the Stanford Daily, a student-run campus newspaper, to generate multiple versions of articles they'd written that were available on the Daily's website. We hypothesized that writers at the Daily would be able to effectively condense their articles such that the newspaper's typical readership (i.e. university students) would be able to easily gather the gist of the article's content. Further, we believed that readers who were allowed to switch between various degrees of content granularity would be better able to recollect an article's key points after reading it under a time constraint than those who were not allowed to do so.

We developed two interfaces: one for readers to read content presented in optionally short, medium, or long versions, and one for authors to upload content reconfigured to be read in this manner.

THE GIST SYSTEM
Gist comprises two interfaces. The first presents articles that have been partitioned into several sections. Each section has a long version (the section as written in its original, unedited form), a short length version that consists of a one to two sentence summary of the entire section, and a medium length version that is of some gradient length between that of the short and long versions. The second interface allows article authors to upload articles that they've adjusted to fit this scheme.

The Reading Interface
When users open an article, they are presented with what on first glance resembles a standard news website. The page's text, however, is significantly abbreviated—sections begin displayed in their short form. Each section is flanked by two buttons at its right side—an up arrow labeled "Shrink" and a down arrow labeled "Expand". Clicking either one of these

Figure 1. Above is the reading interface for an article formatted by its author. Sections are distinguished by a thin grey line. Each section has two buttons beside it that allow a reader to either shrink or expand its text.

buttons switches the section text to that of one of the short, medium, or long versions of that section.

The Upload Interface

Article authors upload their work via the upload interface. The author is presented first with a form that allows them to provide meta information about the article they want to upload—its title, the author's name, a primary image, etc. Below this form are three aligned rich text editors labeled "Short", "Medium", and "Long", footed by buttons that allow the author to add another section or submit what they've filled in. Clicking the "Add section" button renders a new row of text editors below the last. Each of the three text editors is captioned with a suggestion on the content of each version of a section. For example, the caption above the "Short" text editor reads "The short version of a section should state the main idea in one sentence, or two, if absolutely necessary."

RELATED WORK

This project is related primarily to work in usability design for content delivery under short time constraints. Research on the methodology behind skim reading that notes its ineffectiveness [4] and the effect obscuring content has on information retrieval [2] [7] motivate research like this.

Though we ultimately decided to have writers at the Daily condense articles for the sake of practicality, we did consider other methods for shortening articles. We considered employing one of several freely available text-summarization API's, but found the summaries insufficient. We looked to past work implementing crowd sourced systems that relied on Amazon's Mechanical Turk [1] [6] [8]. Given that we had easy access to writers—they were fellow students—and that none of us had any prior experience with Mechanical Turk, we decided that, beyond the knowledge of the article granted by the presence of the article's author, having writers at the

Daily condense their articles was our quickest and most effective option.

EVALUATION

Four authors and twenty-four readers were recruited to evaluate the Gist system. The authors reformatted a total of seven articles, six of which were used in the reading interface evaluation. Most of the authors were able to successfully upload their partitioned articles via the upload interface. Authors who uploaded their articles after the upload interface was adjusted to take its present form were able to upload articles on their own with only an introductory email that explained the Gist system and linked them to the upload interface to guide them.

Once the articles were uploaded, twenty-four participants volunteered to help test the Gist system's effectiveness. We implemented a within-subjects design described below.

All of the four authors and twenty-four readers were college students, save one recent graduate (age 23). Readers spanned eighteen majors. Eight of our twenty-four readers were female and the other sixteen were male. Two of our four authors were female and the other two were male.

Study design

Each of the twenty-four participants read two articles, once in a typical online news reader format (i.e. the reading interface with each section expanded to its long form and the buttons and section dividers hidden), and once with the Gist system's functionality enabled. Readers were given three minutes to read an article approximately eight hundred to nine hundred words long, then were given an eleven question quiz about the article. They would then read a new article in the alternate form and answer eleven more questions about it. Twelve participants were given an article in its standard form before being given an article in its Gist-enabled form. The other twelve participants read and were quizzed on a partitioned article before reading an article its standard format.

We logged the time it took readers to finish reading each article, their viewport upon scrolling, and, in the Gist-enabled articles, which sections they shrank and expanded.
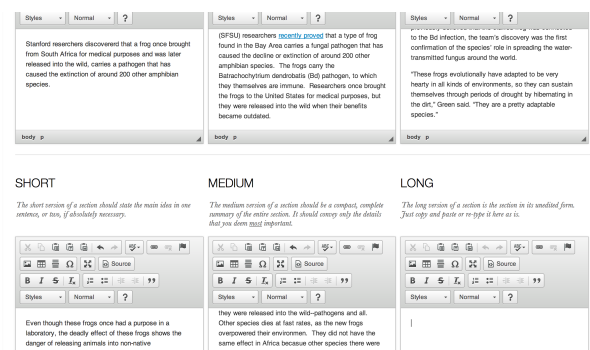

Figure 2. The upload interface allowed authors to partition their articles and draft short, medium, and long versions of each section. Recommendations on the content of each section were placed directly above each text editor.
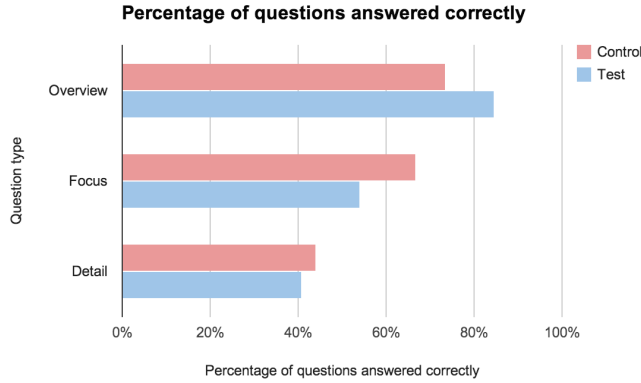
Figure 3. Percentage of questions answered correctly across three question categories

Figure 4. Percentage of questions on news articles answered correctly

Figure 5. Percentage of questions on non-news articles answered correctly

Each of the eleven question quizzes followed what we deemed an appropriately indicative format. These quizzes were written independently of the article's shortened versions—each of us only wrote quizzes for articles we hadn't seen in their edited form, and were administered verbally. Readers were not allowed to look back at the article they were being quizzed on. The task design borrowed from [2] Quiz tasks fell into one of three categories: overview, focus, and detail.

- Overview tasks required subjects to answer general questions that covered the entire range of the article (e.g. "What was this article about?" or "What argument does the author make?")

- Focus tasks required subjects to answer questions about the article's central ideas (e.g. "What is the universitys biggest reason for not implementing this service?" or "According to the author, what is more empowering than 'finding your passion'?")

- Detail questions tested recall of very specific pieces of information contained in the article (e.g. "How often does the Committee on Libraries meet per year?" or "How many South African clawed frogs were tested?")

   Each quiz consisted of three overview, three focus, and five detail questions.

RESULTS

Twenty-four participants read two articles each, once in a standard news reading interface and once in the reading interface implemented as a part of the Gist system.

We found that subjects performed better on overview tasks, significantly worse on focus tasks, and about as well on detail tasks in the test condition compared to performance in the control condition. A paired t-test that compared overview scores on the control condition (mean=2.2083, std. dev=0.6594) to scores on the test condition (mean=2.54167, std. dev=0.6580) neared significance (t(23)=-1.6213, p=0.1186, d=0.3309). A paired t-test that compared focus scores on the control condition (mean=2,
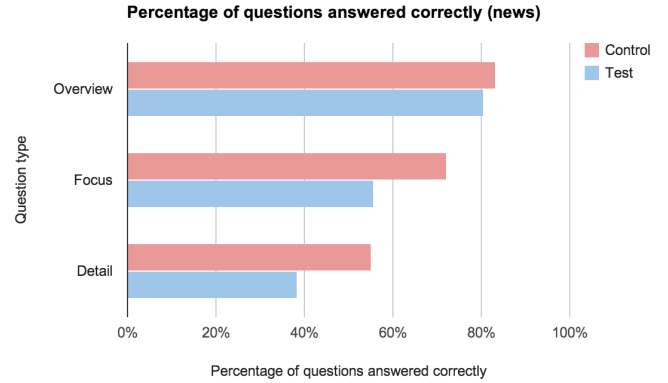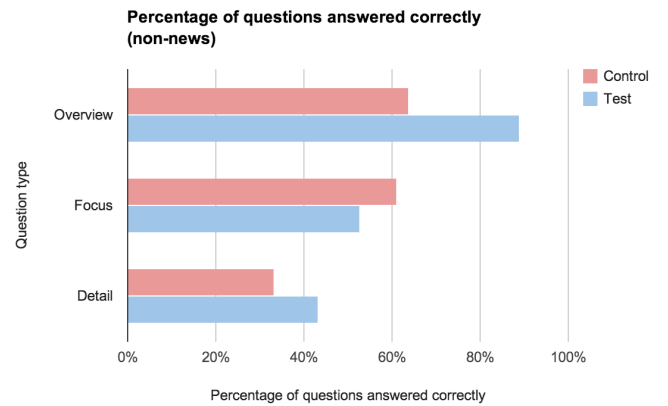
std. dev=0.6594) to scores on the test condition (mean=1.625, std. dev=0.8242) was significant (t(23)=1.8946, p=0.07078, d=0.3867). A paired t-test that compared detail scores on the control condition (mean=2.2083, std. dev=1.3825) to scores on the test condition (mean=2.04167, std. dev=1.4590) was insignificant (t(23)=0.4239, p=0.6756, d=0.08653).

Interestingly enough, and perhaps counter-intuitively, additional chi-squared tests showed that the subjects performed *better* when using the test interface to read non-news articles (three of the six articles used) than they did when using a standard interface (Overview (better): p=0.02642, Focus (worse): p=0.6341, Detail (better): p=0.3478). Conversely, the test interface performed worse when used to read straight news articles (Overview (worse): p=1.0, Focus (worse): p=0.2199, Detail (worse): p=0.09959).

We ran a t-test for differences between article authors and found a significant difference between the two authors in the non-news category (Overview: p=0.09814, Focus: p=0.04983). However, a very low sample size at the author by author resolution (n=8) may explain this apparent significance. A similar significance appeared between the writers of news articles (Detail: p=0.0175). Again, a very low sample size (n=8) may explain this apparent significance.

DISCUSSION AND FUTURE WORK

One author said that, given that he'd known he would eventually have to break down and edit his article to fit this format, he would pay more attention to just how each point in his argument was laid out and how clearly each idea was articulated, so that concepts could be more easily distilled into a short version.

On the editing process, this author said the ideas were "clear in my mind, especially because I wrote it". He added that he didn't have to think very hard about what the main ideas were on the second go round. Another author reported, however, that, since the article she edited down wasn't very complicated—it was a straight news article about a pathogen introduced to the Bay Area by a species of South African frogs—"someone else could have done it just as well".

Authors reported spending fifteen to twenty minutes condensing each article. One author claimed that she was more concerned with completing the editing task than she was with the readability for the end reader. She did, however, end up revising breaks in her article she'd put in place when she wrote it to better encapsulate the article's main ideas.

Both of these authors said that they felt the format was more appropriate for news stories that could be delivered as straight facts. As it turned out, the format appears to be better for articulating opinionated arguments than it is for relaying news. Both the authors and the readers seemed to agree that reading non-news articles in this format seemed to break the article's flow—one author claimed that few, if any, jokes in his satirical tribute to a fallen public figure made it into the short version of a section. In short, this format doesn't aim to render beautiful prose more beautiful, and doesn't lend itself well to writing that derives most of its value from aesthetic or highly stylized language.

It does, however, seem to bolster written arguments' transparency and permanence. One reader said that, though she preferred reading in the standard format, she found points made in the short versions of sections easier to remember.

This sentiment's prevalence presents a very interesting challenge: people generally did not find reading with the Gist system's interface enjoyable when compared to the experience of reading in the long form, but they recognized that the interface had its use case. Further tests could try articles of greater length and complexity and see whether readers still preferred reading in the standard interface or the Gist interface.

A few test subjects would skip over the interface altogether, preferring to begin a session by expanding each section to its long form before beginning to read anything. Others would expand an article to its long form and, on seeing the section expand, would immediately shrink it down to its medium length. We consider this second user a further indication of the existence our intended set—people who are hesitant to jump into lengthy bodies of text.

One concern raised was the visibility of the short versions of sections in the longer versions. Users who expanded a section were surprised to find that the sentence or two that made up the short section were lost in the medium and long forms—main ideas were buried under text. We'd discussed highlighting words that appeared in the short section in the medium section and words that appeared in the medium section in the long section, but it was simply never implemented. It would be interesting to see what affect that would have on the readability of article content and the overall user experience of reading with the Gist interface.

CONCLUSION

In this paper, we introduced a model for displaying online news content in which a reader is given an article in an optionally condensed form in order to improve comprehension when reading under a time constraint. The reader can toggle the length of an article to read each section and topic in more or less detail. We found that this system's approach had the greatest effect on comprehension of non-news articles that presented subjective arguments, though readers still generally preferred reading these types of articles in a standard format. Our results suggest potential avenues for future work in the area of interfaces for rapid, impactful information delivery and interfaces that provide content at optionally variable resolution.

REFERENCES

1. Bernstein, M. S., L. G. M. R. C. H. B. A. M. S. K. D. R. P. K. Soylent: a word processor with a crowd inside. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, ACM (October 2010), 313–322.

2. Buchanan, G., O. T. Improving skim reading for document triage. In *Proceedings of the second international symposium on Information interaction in context*, ACM (October 2008), 83–88.

3. Das, D., and Martins, A. F. A Survey on Automatic Text Summarization.

4. Duggan, G. B., . P. S. J. How much do we understand when skim reading? In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, ACM (April 2006), 730–735.

5. Hewitt, J., B. C. P. V. Scan rate: A new metric for the analysis of reading behaviors in asynchronous computer conferencing environments, December 2007.

6. Kittur, A., C. E. H. S. B. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM (April 2008), 453–456.

7. Lam, H., B. P. Summary thumbnails: readable overviews for small screen web browsers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (April 2005), 681–690.

8. Snow, R., O. B. J. D. N. A. Y. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks, October 2008.