Auto-Highlighter: Identifying Salient Sentences in Text

Jessica Zeitz Self, Rebecca Zeitz, Chris North
Department of Computer Science
Virginia Tech
Blacksburg, VA, USA
{jzself, razeitz, north}@cs.vt.edu

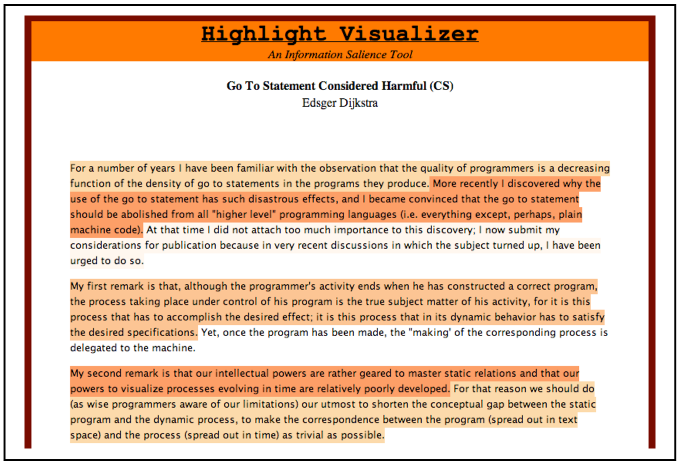
Alan L. Breitler
Simulation and Systems Division
DDL OMNI Engineering LLC
Virginia Beach, VA, USA
alan.breitler@ddlomni.com

***Abstract*—To help analysts sift through large numbers of documents, we suggest an auto-highlighting system that computationally identifies the topmost salient sentences in each document as a form of summary and rapid comprehension aid. But what is it that makes a sentence or group of sentences important in the context of a document? Can we understand human thought processes in document comprehension to improve computerized selection of salient sentences? We conducted a user study to gather data about the types of sentences people highlight when reading and comprehending text. Users read purely textual documents, highlighted important sentences, and then explained why they selected those particular sentences. Our study focuses not only on the comparison between expert and non-expert users for different document types, but also the comparison between users and common algorithmic metrics for sentence selection. We provide a user-defined categorical approach to describing the variations in the types of highlighted sentences as well as insight concerning rhetoric and language that could strengthen future algorithms.**

Keywords—auto-summarization comprehension; user study; computer-aided

I. Introduction

With the rise of big data comes the increasing need for methods to help people rapidly comprehend it. Industrial and government agencies receive and collect more data than ever before and now need ways to analyze it more efficiently. This data takes many forms including textual data from documents such as emails, transcriptions, and reports. Textual data, which often masks its wealth of information through its form, presents the ever present issue of comprehension. Document comprehension is often a challenging task, one that has proven so difficult that humans employ an array of techniques such as annotating (highlighting, underlining, etc.), summarizing, and rereading in order to improve understanding. Analysts must routinely sift through numerous documents and determine if and how they are relevant or important to the problem at hand. Given so many documents, an analyst does not have time to read each one in detail. She has to scan at a high level to decide what is worth her time to read. Analysts need new tools that reduce the amount of data to be reviewed and provide an overview to more efficiently assess the documents. Imagine automatically highlighting important information in the documents, or reducing a digital stack of documents down to one-paragraph summaries that are automatically constructed through sentences taken directly from the documents.



Highlight Visualizer
An Information Salience Tool

Go To Statement Considered Harmful (CS)
Edsger Dijkstra

For a number of years I have been familiar with the observation that the quality of programmers is a decreasing function of the density of go to statements in the programs they produce. More recently I discovered why the use of the go to statement has such disastrous effects, and I became convinced that the go to statement should be abolished from all "higher level" programming languages (i.e. everything except, perhaps, plain machine code). At that time I did not attach too much importance to this discovery; I now submit my considerations for publication because in very recent discussions in which the subject turned up, I have been urged to do so.

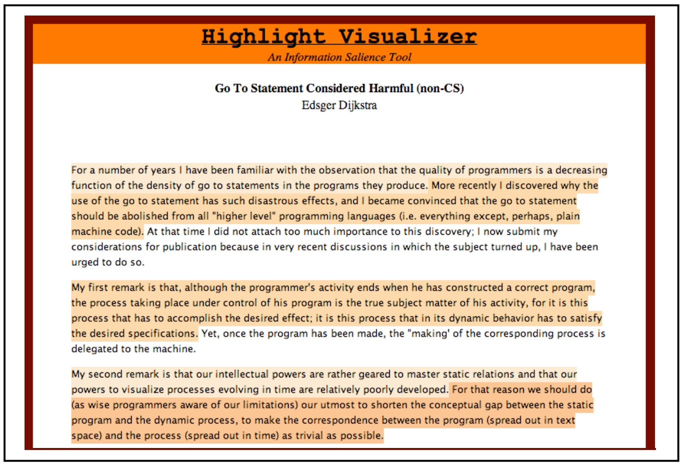
My first remark is that, although the programmer's activity ends when he has constructed a correct program, the process taking place under control of his program is the true subject matter of his activity, for it is this process that has to accomplish the desired effect; it is this process that in its dynamic behavior has to satisfy the desired specifications. Yet, once the program has been made, the "making' of the corresponding process is delegated to the machine.

My second remark is that our intellectual powers are rather geared to master static relations and that our powers to visualize processes evolving in time are relatively poorly developed. For that reason we should do (as wise programmers aware of our limitations) our utmost to shorten the conceptual gap between the static program and the dynamic process, to make the correspondence between the program (spread out in text space) and the process (spread out in time) as trivial as possible.

Fig. 1. Highlight Visualizer is displaying the first three paragraphs of “Go To Statement Considered Harmful” [9]. The highlights depict the important sentences selected by multiple readers; the darker the orange the more readers selected the sentence. In this case, the readers are technical experts.

We suggest an automated approach that selects the topmost salient sentences from a document, thus reducing the textual information. This provides the analyst with a short summary of the document that includes entire sentences taken straight from the text. Entire sentences retain the properties of natural text, allowing the analyst to easily read, comprehend and evaluate its possible importance.

By selecting entire sentences, the results can be used in or out of the context of the document. Selected sentences from the document are combined to create a standalone paragraph



Highlight Visualizer
An Information Salience Tool

Go To Statement Considered Harmful (non-CS)
Edsger Dijkstra

For a number of years I have been familiar with the observation that the quality of programmers is a decreasing function of the density of go to statements in the programs they produce. More recently I discovered why the use of the go to statement has such disastrous effects, and I became convinced that the go to statement should be abolished from all "higher level" programming languages (i.e. everything except, perhaps, plain machine code). At that time I did not attach too much importance to this discovery; I now submit my considerations for publication because in very recent discussions in which the subject turned up, I have been urged to do so.

My first remark is that, although the programmer's activity ends when he has constructed a correct program, the process taking place under control of his program is the true subject matter of his activity, for it is this process that has to accomplish the desired effect; it is this process that in its dynamic behavior has to satisfy the desired specifications. Yet, once the program has been made, the "making' of the corresponding process is delegated to the machine.

My second remark is that our intellectual powers are rather geared to master static relations and that our powers to visualize processes evolving in time are relatively poorly developed. For that reason we should do (as wise programmers aware of our limitations) our utmost to shorten the conceptual gap between the static program and the dynamic process, to make the correspondence between the program (spread out in text space) and the process (spread out in time) as trivial as possible.

Fig. 2. Highlight Visualizer is displaying an excerpt from “Go To Statement Considered Harmful.” The readers for this visualization are non-experts.

serving as a summary. Within the context of the document, the sentences can be visually highlighted, providing cues to important points and aiding in comprehension. These two methods, in context and out of context, can be used together in user interaces for document analytics systems, allowing the analyst to interact within a filtering tool by progressively adding or reducing the amount of visible sentences. The original summary sentences provide valuable anchors within the document, from which the surrounding context naturally extends by exposing additional sentences or the full text.

To inform the design of algorithms for automatic highlighting, we look to human behavior. The idea of auto-highlighting leads to several important research questions that we investigate in this paper:

- Are human highlighted sentences representative of human synthesized summaries? When summarizing, how much do humans rely on sentences they previously highlighted? This question is necessary to investigate the claim that salient sentences will provide good summaries. (Q1)
- Which sentences in a document do humans deem salient? Why? This question seeks a basis for defining sentence salience. (Q2)
- Are there differences between experts and non-experts when highlighting and summarizing a document? This question investigates the importance of domain knowledge to salience selection. (Q3)
- How closely can simple algorithm heuristics mimic the human selection of salient sentences? This question investigates the feasability of simple computational methods. (Q4)

To answer these questions, we conducted a user study to discover metrics behind the complex cognitive process humans go through when highlighting and summarizing a document. We aimed to compare user-generated data against algorithm-generated data about extraction-based document summaries. We developed an algorithm to select sentences based on four textual metrics for salience. We designed a user interface that highlights sentences based on their salience scores. The interface provides a visual representation of the importance of each sentence in relation to the document, as seen in Fig. 1.

II. Related Work

Auto-summarization of text includes two types of methods: extraction and abstraction. Both techniques have been studied, focusing on the usage of multiple metrics within algorithms. Previous work has found that intelligent summarizers do outperform summarizers that randomly select sentences. Many algorithms have then been tested against human-generated summaries with user studies [1–4]. [1] performed a study comparing the selection and rank of 20 sentences selected by humans and by five algorithmic methods. Most of the analysis consisted of computational statistics about agreement and disagreement between the humans and the algorithms. The authors briefly discussed the role of topic sentences within the selected sentences, however this is only starting to delve into the relation between selected sentences and sentence type. The authors of [4] performed a similar study that not only compared auto-summaries to human-generated reference summaries, but also had humans rate the auto-summaries.

There are multiple summary evaluation methods [4–7], which are largely based on the context of the documents and the reasons and criteria for the summary creation, even when trying to create generic summaries. The evaluation of auto-generated summaries and the automatic summary generator itself is challenging given that there is low agreement among users when determining what sentences should be included in a summary [1], [7], [8]. It has been suggested that due to this low agreement, multiple reference summaries should be used when evaluating auto-generated summaries [8]. One study found that the length of a summary is independent of the length of the document, and as such, using a constant summary length is better than adjusting the summary length to the document length [5]. Given the difficult evaluation process, studies have been conducted to gather insight into the types of metrics and features that can be used when automatically generating summaries [1], [2], [3], [5]. However, there is more work to be done in determining which features of the document can be exploited when developing a document auto-summarizer.

III. Experiment

We conducted an experiment to learn about how and why users highlight sentences in documents, and how that compares to algorithmic approaches. We used a 2x2 counter-balanced factorial design where all participants in 2 groups read two documents, with half reading one document first and half reading the other document first.

The study focused on two independent variables which guided our design. The first variable was *document type*, and we used one technical (computer science related) article and one general or non-technical article. We selected essays that attempted to argue a point, and were less than 2 pages in length. The technical document was “Go To Statement Considered Harmful” by Edsger Dijkstra [9], and the non-technical document was “Time Wars” by Mark Fisher [10].

The second independent variable was the *participants’ area of expertise*. We had 40 participants; 20 computer science majors and 20 non-computer science majors. Non-computer science participants were from 13 different disciplines, including psychology, which made up 6 of the participants, and geography, which made up 3 of the participants. All participants were undergraduate or graduate students. The average age for the computer science group was 23.75. This group consisted of 6 females and 14 males, where 11 were undergraduate students and 9 were graduate students. The average age for the non-computer science group was 19.65. This group consisted of 14 females and 6 males, all of whom were undergraduate students. Computer Science participants were considered experts for only the technical document, whereas no participant was considered an expert for the non-technical document. No participant had previously read either document.

Our four dependent variables (highlights, reasons for highlights, summaries, and how highlights relate to summaries) directed the tasks we chose for the study. Participants were first

TABLE I. CONNECTING HIGHLIGHTS AND SUMMARIES

Category	Go To Statement		Time Wars	
	Non-technical	*Technical*	*Non-technical*	*Technical*
Total #	47	47	80	80
Total sentences not highlighted by anyone	22	21	33	29
Total sentences highlighted by at least 1 participant	25	26	47	51

asked to read a document and highlight the five "most important sentences that will best help you summarize the document". Second, the participants were asked to summarize that document. We designed the study so that the highlighting was embedded in a larger comprehension task. Thus the participants had a goal in mind when selecting their highlights. Finally, the participants were asked to specifically explain why they highlighted the five particular sentences. Participants were then asked to repeat these three steps with the second document. The study concluded with a questionnaire asking participants whether they had read the documents previously, to what extent their highlights helped when summarizing, and to provide an explanation about their process when reading, highlighting and summarizing. None of the participants had read "Time Wars" before the study. While none of the non-technical participants had read "Go To Statement Considered Harmful" prior to the study, there were two technical participants who were familiar with the content of the document without having actually read it. One had skimmed the document before, and another had read a summary of the document. All of the participants noted that the highlights helped, although how much and for what purpose the highlights helped varied. However, 35 out of 40 participants used highlights to identify main points in the documents or to serve as general points of reference (location reminders or content reminders). Notably, 13 out of 20 participants in each group, for a total of 26 out of 40, read some or all of the document before highlighting, even though they could unhighlight. Participants gave reasons for this method including that they could only highlight five sentences and that they wanted to get a better understanding of the document and its main points first.

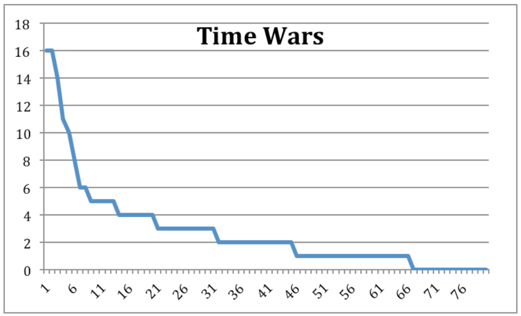


Fig. 4. The power-law distribution for "Time Wars" showing ranking of sentences (X-axis) based on number of participant highlights (Y-axis).

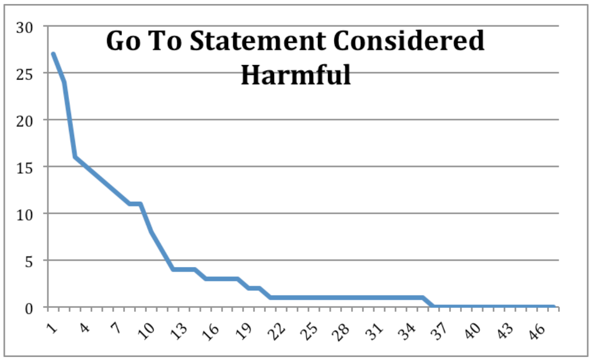


Fig. 3. The power-law distribution for "Go To Statement Considered Harmful" showing ranking of sentences (X-axis) based on number of participant highlights (Y-axis).

IV. FINDINGS

As instructed, most participants highlighted 5 sentences. Only 7 and 11 participants highlighted more or less than 5 in "Go To Statement Considered Harmful" and "Time Wars" respectively. "Go To Statement Considered Harmful" contains a total of 47 sentences, 22 of which were not highlighted by any non-technical participants and 21 of which were not highlighted by any technical participants. A total of 25 sentences were highlighted by at least one non-technical participant and a total of 26 sentences highlighted by at least one technical participant. "Time Wars" contains a total of 80 sentences; 33 were not highlighted by any non-technical participant and 29 were not highlighted by any technical participant. There were a total of 47 sentences highlighted by at least one non-technical participant and a total of 51 sentences selected by at least one technical participant. Overall, there were 25 sentences not highlighted by any participant in "Go To Statement Considered Harmful" and 27 sentences not highlighted by any participant in "Time Wars." (See Table 1)

As seen in Figs. 3 and 4, the highlighted sentences form a power-law distribution such that there are only a few highly rated sentences and most other sentences were highlighted by only 1 or 2 participants.

A. *Characterization of Summaries*

We analyzed each participant's summary, comparing it to his or her five highlighted sentences and looking for direct quotes and sentence paraphrasing. We counted the number of ideas within each summary and then noted whether the idea could be traced back to a highlighted sentence or another sentence in the document. If either trace is possible, we say it was synthesized information. The following is a summary written by a non-technical participant for "Go To Statement Considered Harmful." The bracketed information relates to the participant's highlighted sentences; double brackets denote information stemming from a sentence in the document, but not one that was highlighted.

> *"[The process that is taking place and being controlled under one's own program is the actual activity], even though [a programmers activity ends when he has made a correct program]. [[Due to the fact that our ability to visualize processes happening over time is poor]], [we*

> *need to make the gap between the static programming and dynamic process more specific]. To make this more suitable we can [add textual and dynamic arrangements to the program]. These concepts may be helpful, however they are [out of the programmers control due to the fact that they are part of the write- up of the program]. In all,the [programmer should make his/her system more manageable and suitable for everyone by adding certain clauses to help progress and process].”*

TABLE II. PROMINENT CATEGORIES OF REASONS FOR HIGHLIGHTS

Category	Go To Statement		Time Wars	
	Non-technical	*Technical*	*Non-technical*	*Technical*
Argument/ main point	13	27	14	11
Supporting evidence	54	45	54	54
Solution	7	0	7	1
Profound statement	n/a	n/a	6	0
Personally resonated	n/a	n/a	0	6
Conclusion	10	6	4	8

This example contains seven key elements; six stemming from the participant’s highlighted sentences and one stemmming from another sentence in the document. Table 3 shows the percentages of ideas stemming from highlighted sentences, other sentences within the document, or synthesized information.

Based on this data, elements within summaries can be traced back to the reader’s highlighted sentences answering Q1. Specifically, summaries are approximately 50% characterized by participants’ top five sentences. We conclude that sentences highlighted by a reader are well-representative of a user-synthesized summary (Q1).

B. User-defined Categories

As people read, certain elements stand out as being more important than other elements. But what characterizes these elements and are there characteristics generally consistent from one person to the next? To answer such questions, we asked participants to explain the reason behind their selection of highlighted sentences. The analysis helped answer Q2.

Through open coding, we condensed the reasons for highlighting into 12 categories for the technical document and 21 categories for the non-technical document. These categories were user-defined and stemmed from the reasons participants gave for highlighting certain sentences. Such categories included: introduction, background information, concept connection, example, paragraph summary, and conclusion. The top two categories for both documents were an *argument/main point* sentence and a *supporting evidence* sentence. Table 2 shows the number of instances where some sentence was labeled with a category by a participant. For example, there were 45 instances where some sentence in “Go To Statement Considered Harmful” was labeled by a technical participant as a *supporting evidence* sentence.

Most highlighted sentences were labeled differently by individual participants. Overlaps occurred across the *argument/main point, supporting evidence,* and *conclusion* categories. The following sentence was labeled as both an *argument/main point* and *supporting evidence.*

> *“The main point is that the values of these indices are outside programmer's control; they are generated (either by the write-up of his program or by the dynamic evolution of the process) whether he wishes or not* [9]*.”*

This sentence contains the actual words *main point* which suggests importance, but the sentence could explain why the author believes ‘go to’ statements are detrimental. Even though participants were not completely consistent with reasoning, the sentence was important enough to highlight.

TABLE III. CONNECTING HIGHLIGHTS AND SUMMARIES

Reference Location	Go To Statement		Time Wars	
	Non-technical	*Technical*	*Non-technical*	*Technical*
Highlighted sentences	51%	50%	44%	59%
Elsewhere in document	10%	8%	41%	32%
Synthesized information	39%	42%	15%	9%

TABLE IV. CORRELATION R VALUES BETWEEN METRICS AND HUMANS

Document	M1	M2	M3	M4	Total
Time Wars	0.542	0.541	0.335	0.612	0.570
Go To Statement	0.288	0.288	0.046	0.398	0.348

The *solution* category produced interesting numbers. Seven non-technical participants labeled some sentence in “Go To Statement Considered Harmful” as a *solution*, however no technical participants used this label. The same is true for “Time Wars”; many more non-technical than technical participants categorized sentences as *solutions*.

The *profound statement* and *personally resonated* categories produced opposite results as seen in Table 2. This sentence from “Time Wars” was labeled as a personally resonating sentence by two technical participants:

> *“The consequence is a strange kind of existential state, in which exhaustion bleeds into insomniac overstimulation (no matter how tired we are, there is still time for one more click) and enjoyment and anxiety co-exist (the urge to check emails, for instance, is both something we must do for work and a libidinal compulsion, a psychoanalytic drive that is never satisfied no matter how many messages we receive)* [10]*.”*

We can speculate that the participants related to “insomniac overstimulation” and having “the urge to check emails” given their involvement in a technical field, Computer Science.

C. Rhetorical Structure

The analysis of categories exposed another finding. Categories stem from the elements of the rhetorical structure of a document. We found that categories chosen by participants strongly correlated with elements such as introduction, main point, supporting evidence, and conclusion. These elements are the focus of readers and writers since they provide a basic structure for organization of a document. Categories based on more formalized rhetorical elements (i.e. main point, supporting evidence) were used more often than other elements when labeling sentences.

This finding suggests that sentences fitting in one of these main rhetorical elements are more likely to be selected by a reader as salient.

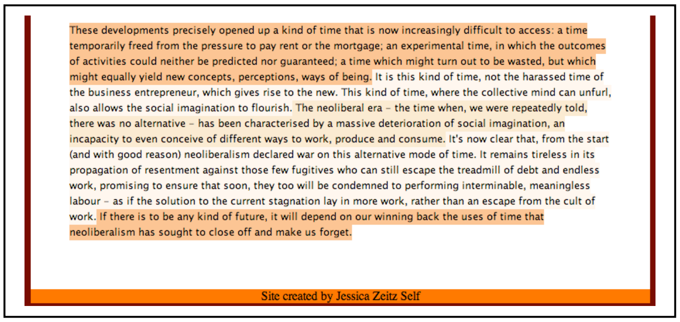
Sentences within these categories also fit rhetorical structure as it pertains to ordering. We found that sentences selected as *introduction* or *background* sentences most often appeared in the first few paragraphs of a document whereas sentences categorized as *conclusion* appeared toward the end. This phenomenon occurred in both the technical document and the non-technical document.

D. Experts versus Non-experts

Variations between experts and non-experts were minimal. In general, experts and non-experts highlighted similar sentences (Q3). For “Go To Statement Considered Harmful” any sentence that was selected by 7 or more participants was considered salient. An average of 7.3 non-technical participants selected a salient sentence for “Time Wars,” compared to an average of 7.8 technical participants. For “Time Wars” any sentence that was selected by 10 or more participants was considered salient. An average of 7.6 non-technical participants selected a salient sentence for “Time Wars,” compared to an average of 5.8 technical participants. Considering the latter document, the percentage of total salient highlights is 57% for non-technical participants and 43% for technical participants. Considering the former document, the percentage of total salient highlights is 48% for non-technical participants and 52% for technical participants.

V. Algorithm

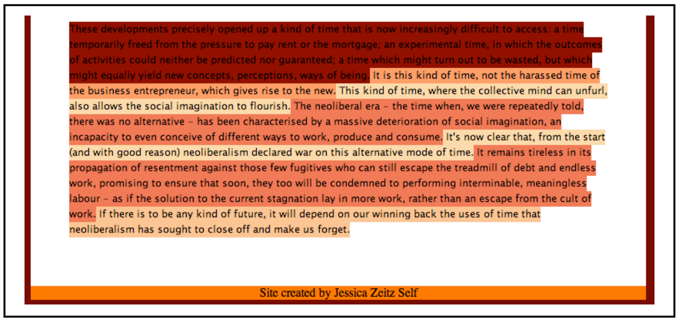
We developed an algorithm that selects salient sentences based on the following four simple and common textual metrics (all exclude stop words). These metrics attempt to identify sentences that are representative of the entire document. There is indication from experimentation in the literature that simple text metrics may be adequate, and in some cases outperform more advanced metrics [11].



These developments precisely opened up a kind of time that is now increasingly difficult to access: a time temporarily freed from the pressure to pay rent or the mortgage; an experimental time, in which the outcomes of activities could neither be predicted nor guaranteed; a time which might turn out to be wasted, but which might equally yield new concepts, perceptions, ways of being. It is this kind of time, not the harassed time of the business entrepreneur, which gives rise to the new. This kind of time, where the collective mind can unfurl, also allows the social imagination to flourish. The neoliberal era – the time when, we were repeatedly told, there was no alternative – has been characterised by a massive deterioration of social imagination, an incapacity to even conceive of different ways to work, produce and consume. It's now clear that, from the start (and with good reason) neoliberalism declared war on this alternative mode of time. It remains tireless in its propagation of resentment against those few fugitives who can still escape the treadmill of debt and endless work, promising to ensure that soon, they too will be condemned to performing interminable, meaningless labour – as if the solution to the current stagnation lay in more work, rather than an escape from the cult of work. If there is to be any kind of future, it will depend on our winning back the uses of time that neoliberalism has sought to close off and make us forget.

Site created by Jessica Zeitz Self

Fig. 5. Highlight Visualizer is displaying the last paragraph of “Time Wars.” It includes data based on all 40 readers.



These developments precisely opened up a kind of time that is now increasingly difficult to access: a time temporarily freed from the pressure to pay rent or the mortgage; an experimental time, in which the outcomes of activities could neither be predicted nor guaranteed; a time which might turn out to be wasted, but which might equally yield new concepts, perceptions, ways of being. It is this kind of time, not the harassed time of the business entrepreneur, which gives rise to the new. This kind of time, where the collective mind can unfurl, also allows the social imagination to flourish. The neoliberal era – the time when, we were repeatedly told, there was no alternative – has been characterised by a massive deterioration of social imagination, an incapacity to even conceive of different ways to work, produce and consume. It's now clear that, from the start (and with good reason) neoliberalism declared war on this alternative mode of time. It remains tireless in its propagation of resentment against those few fugitives who can still escape the treadmill of debt and endless work, promising to ensure that soon, they too will be condemned to performing interminable, meaningless labour – as if the solution to the current stagnation lay in more work, rather than an escape from the cult of work. If there is to be any kind of future, it will depend on our winning back the uses of time that neoliberalism has sought to close off and make us forget.

Site created by Jessica Zeitz Self

Fig. 6. This figure is displaying the last paragraph of “Time Wars.” The Highlight Visualizer is using the algorithm salience scores. The red highlights depict the most salient sentences selected by the algorithm.

M1: The sum of document word frequency counts for each significant word in the sentence.
M2: The sum of the number of different sentences in which each significant word in the sentence appears.
M3: The sum of word frequencies within a 6-sentence window surrounding the sentence as a measure of local relevance.
M4: The total number of n-grams shared with other sentences (n=1..5) as a measure of similarity.

We summed all four metrics to create a total salience score for each sentence.

VI. User and Algorithm Comparison

We compared the user-selected and algorithm-selected salient and non-salient sentences to answer Q4. See Figs. 5 and 6 for visualized data. Table 4 contains the correlation values between the number of users selecting a sentence to the alrogithm sentence salience scores. The total column values correlate the number of users and the four metrics added together. Even though the correlation values are not high, the results show promise that with the correct weighted metrics, an algorithm can supply a user with a representative extraction-based document summary. In general, the metrics performed better for the non-technical Time Wars document. Metric M4, based on n-grams, performed the best overall. M3, which attempted to find localized relevance, failed. The comparison between M1 and M2 indicates that it is not critical whether frequencies are counted based on words or unique sentences.

The results shown in Table 5 are based on selected the top 10 salient sentences in each document, according to the human raters and the algoirthm. However, in Time Wars there was a 5 way tie in the human-raters scores, giving the top 13 salient sentences. For the human raters, these are the sentences the received the most votes from all participants. For the algorithm, these are the sentences the received the highest total salience score. Based on the summary lengths of 10 and 13 sentences respectively for “Go To Statement” and “Time Wars,” the humans and algorithm agreed upon 5/10 50% and 6/13 46% salient sentences. Thus, the algorithm correctly identified about half of the top salient sentences for automatic highlighting, and successfully eliminated about three quarters of sentences as correctly non-salient.

Through analysis, we found that sentence length based on word count is somewhat correlated with whether the sentence was highlighted by users. Users tended to highlight lengthier sentences as opposed to shorter sentences. As a result, normalizing the four metrics by sentence length reduced their correlative power. Thus, we recommend not using length normalized metrics.

As an alternative metric, we also calculated a fifth metric similar to *tf-idf* called term "term frequency / inverse sentence frequence" which measured the uniqueness of each sentence by down-weighting terms that occurred frequently in other sentences. When correlated with the human scores the r value was very low and negative, indicating that users tended to not pick unique, unusual, or odd-ball sentences. This along with the results of the M1-4 metrics help us understand what "salience" means to users. They tend to highlight sentences that are more representative of the document as a whole instead of sentences that are unique.

We conclude that sentences selected by the algorithm can adequately represent a generic summary of the document. Since humans utilize outside knowledge, the algorithm cannot exactly replicate sentences selected by a reader. However, incorporating our other findings will improve and strengthen extraction-based summarization algorithms to more closely mimic human selection.

TABLE V. HUMAN & ALGORITHM AGREEMENT/DISAGREEMENT

Document	True Positive	False Positive	False Negative	True Negative	Total
Time Wars	6	7	7	60	80
Go To Statement	5	5	5	32	47

VII. CONCLUSIONS

This study provides valuable information about human-selected salient sentences, human-generated summaries, and the relationships between the two. Humans exploit rhetorical structure knowledge to pinpoint salient sentences in a document and then use these salient sentences to formulate a summary. Both experts and non-experts of a document employ these methods. Our current algorithm using simple text metrics does a fair job mimicking this human selection. The output is natural to comprehend given it is comprised of complete sentences that can be used as a short summary or as visual highlights in the context of the full text. The findings indicate that such extraction-based summaries composed of salient sentences are well-representative of abstraction-based human-synthesized summaries.

Our Auto-Highlighter technique could be integrated into many text analysis and visualization tools to provide condensed overviews of many documents and support rapid document review by analysts. An interesting possible extension would be to add interactivity, such that the auto-highlighting could respond and adjust to user highlights or other user input, by updating its internal weighting scheme (e.g. [12]).

To further develop algorithm effectiveness, we suggest augmenting these techniques with strategies similar to those used by humans. If the algorithm utilizes basic rhetorical structure (i.e. sentences at the end of the first paragraph) and categorical signals such as main points, it can more efficiently predict salient sentences that would match a human's selection. Future work concerning how additional rhetoric and language concepts can improve algorithmic support for comprehension and will lead us closer to effectively managing big data.

ACKNOWLEDGMENT

The authors would like to acknowledge and thank Paul Abney and John Rice for their support during this research. This research was sponsored by a Small Business Innovation Research (SBIR)(Phase 1) project funded by the Office of the Secretary of Defense and administered by the Naval Research Laboratory (OSD11-TD1).

REFERENCES

[1] G. J. Rath, A. Resnick, and T. . Savage, "The Formation of Abstracts By the Selection of Sentences Part I. Sentence Selection By Men and Machines," *Journal of the American Society for Information Science and Technology*, vol. 12, no. 2, pp. 139–141, 1961.

[2] D. Radev, S. Teufel, and H. Saggion, "Evaluation challenges in large-scale document summarization," *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, vol. 1, pp. 375–382, 2003.

[3] R. Mihalcea and P. Tarau, "A language independent algorithm for single and multiple document summarization," *Proceedings of IJCNLP*, 2005.

[4] C. Hori, T. Hirao, and H. Isozaki, "Evaluation measures considering sentence concatenation for automatic summarization by sentence or word extraction," *Proceedings of the ACL*, pp. 82–88, 2004.

[5] J. Goldstein, J. Carbonell, and M. Kantrowitzt, "Multi-Document Summarization By Sentence Extraction," *Proceedings of the 2000 NAACL-ANLP Workshop on Automatic summarization*, vol. 4, pp. 40–48, 2000.

[6] H. Jing, R. Barzilay, K. McKeown, and M. Elhadad, "Summarization evaluation methods: Experiments and analysis," *Proceedings of the AAAI Symposium on Intelligent Summarization*, pp. 51–59, 1998.

[7] I. Mani, "Summarization evaluation: An overview," in *Proceedings of the NAACL 2001 Workshop on Automatic Summarization*, 2001.

[8] C. Lin, "Looking for a few good metrics: Automatic summarization evaluation—how many samples are enough," *Proceedings of the NTCIR Workshop*, 2004.

[9] E. W. Dijkstra, "Letters to the editor: go to statement considered harmful," *Communications of the ACM*, vol. 11, no. 3, pp. 147–148, Mar. 1968.

[10] M. Fisher, "Time Wars," *Gonzo (circus)*, no. 110, 2012.

[11] X. Wang and A. Kabán, "Model-based estimation of word saliency in text," *Discovery Science*, 2006.

[12] A. Endert, P. Fiaux, and C. North, "Semantic Interaction for Sensemaking: Inferring Analytical Reasoning for Model Steering," *IEEE Conference on Visual Analytics Science and Technology*, 2012.