Skimming for the Visually Impaired

Leah A Judd
Division of Science and Mathematics
University of Minnesota, Morris
Morris, Minnesota, USA 56267
juddx097@morris.umn.edu

ABSTRACT

This paper describes studies conducted to develop and test designs and interfaces made for accessible auditory skimming for visually impaired users. Researchers observed how sighted users skim, and used the findings to develop design guidelines. We discuss the natural language parsing and machine learning approach used by the research group at Stony Brook to create a variable and flexible skimming experience. Another group at the University of British Columbia designed an app for an eyes-reduced skimming experience. Each study brings visually impaired users closer to browsing text as quickly and easily as sighted users.

1. INTRODUCTION

Getting through text quickly is an important ability to have, and skimming is one way to do that without sacrificing comprehension. Sighted users can gather information at a glance, scan over text, and slow down to read information that seems important. Visually impaired users depend on text-to-speech tools to get through text. Having text narrated aloud is time consuming and requires constant focus to follow along. Visually impaired users who want to go through text rapidly are limited in options. Making the text-to-speech narration faster is mentally straining since users need to filter out unimportant information as it is being read out loud. Skipping over full sentences means a lot of information is lost. Text-to-speech on its own is not suited for skimming.

The studies described in this paper worked towards an audio interface to simulate selective reading. This paper summarizes a preliminary study of ad-hoc audio skimming, the process of automating summaries, and the improvements made on interface for skimming. Researchers at Stony Brook University have been developing assistive technology for visually impaired users to effectively skim through written media. Additionally, it summarizes a similar skimming application for situational visual impairments.

2. WHAT IT TAKES TO SKIM

It is good to understand the advantages sighted users have when quickly browsing through written text. For sighted users, there are a variety of tools and techniques available to make skimming possible. In the study conducted by Machulla et al. [9], sighted participants were asked to describe aloud their methods when skimming over different materials, such as scientific articles and textbooks chapters. The participants were recorded skimming over the materials. The sighted participants were given five minutes for each piece of text to provide an overview of the contents. The recordings were analysed, and the results showed the sighted participants used spatially-coded information to help them skim. Spatial-coding refers to the two dimensional spacing and placement of text.

Some skimming techniques use the ability to view the text as a whole, some are to search for specific information, and some use the ability to explore non-linearly, like jumping from an introduction to a conclusion. Sighted users can look at macro-structures, such as paragraphs, tables, and figures. They can also notice micro-structures, like bold, italicised, or colored words, lists, and bullet points. There is keyword spotting, where sighted users can brush over the text searching for specific words or phrases. Finally, there is selective reading, where the users can quickly go through the text, and slow down when the text seems important.

For blind users, the most common way of “skimming” is to listen to an entire document at an increased speed and mentally filter out unimportant information. That is time consuming, requires focus, and is mentally straining. If a user loses focus they can miss important information and have to go back and listen to the text all over again. Some strategies do exist, for example, using a table of contents to link to sections or pages, or the ability to skip between section headers. However, they are not available for many kinds of reading materials.

3. ANALYSIS AND STUDY

A team at Stony Brook University set up a study, described in Ahmed et al. [2], which:

1. Helped identify the type of skimming that can be useful in screen reading main content in web pages
2. Led to the development of a usable interface for accessible online skimming
3. Demonstrated the utility of the accessible skimming interface in two realistic use scenarios
4. Identified automatic summarization techniques that could “closely” approximate skimming methods used by sighted people

UMM CSci Senior Seminar Conference, October 2020 Morris, MN.

To start, the researchers had 12 sighted participants summarize 6 articles, 2 articles each. The 6 articles got 4 summaries that way. The articles themselves were 5-6 paragraphs long. The guidelines for summarizing the articles were:

- For each sentence to be summarized individually
- The summary had to include only the words in a sentence in the order that the words appeared
- The summarized length, or, the number of words in a sentence, should be no longer than one third of the length of the original sentence
- The summaries should be as informative as possible

The summaries were then analysed.

The researchers found that the original text was composed of 31% nouns, 15% verbs, 13% prepositions, 9% adjectives, 4% adverbs and 28% other part of speech. In comparisons, the summaries were made up of 54% nouns, 12% verbs, 11% adjectives, 11% adverbs, 7% prepositions, and 5% other parts of speech.

With nouns taking a huge percentage, it was clear that nouns are an informative part of speech for summaries.

Next, the team condensed the 4 summaries for each article into one summary for each article and called them the “Gold Standard” summary. The Gold Standard Summary was created by picking words that at least 2 of the participants had chosen for their summaries.

The researchers then compressed the summaries even further into different types. The summaries types were labeled A, B, and C. Summary A contained only nouns, summary B contained nouns and prepositions, and summary C was the Gold Standard. Table 1 has an example of the summary types.

Once the preliminary work of creating summaries was completed, the user study was performed. The study was split into a listening-and-comprehension portion, and a search scenario. The participants in the user study were 20 blind users, ranging in age from their 20s to their 60s, all at least comfortable with using a computer and the “Job Access With Speech” (JAWS) [7] screen reader.

3.1 Comprehension

For the listening-and-comprehension part of the study, the users listened to 4 different articles, one in each style of summary and the full text. At the end of each article, the users were given 10 questions to answer. The questions followed distribution of the parts of speech: they had more questions about nouns and verbs than adjective/adverbs. Some examples of the questions can be seen in Table 2.

The results of the comprehension portion of the study showed that comprehension fell as the amount of words in the summaries dropped. The question “what is the article about?” was answered with 100% accuracy for the full text and Gold Summary C. For summaries A (noun-only) and B (noun and prepositions) the question was answered with 80% and 90% accuracy respectively.

When asked about their experience afterwards, users felt that important information was lost in the sparser summaries.

3.2 Search Scenario

The next part of the study, the search scenario, only used the Gold Standard summary and the full text. The participants were given a question before going through an article and were asked to find the answer. The participants went through one article with its full text, and then another article using skimming. When skimming through the given article, the participants used a keyboard shortcut to switch between the Gold Standard summary and the full text. When switching, the screen reader read out the last word that was in both the full text and the summary.

The researchers recorded the time it took users to reach the answer and the time taken to answer the question. There was a significant time improvement from reading to skimming, with the time taken to reach the answers being 1.9 times faster with skimming. Answering the questions was 1.6 times faster using skimming. This demonstrated how much of a time saver skimming could be for users.

4. GENERATING SUMMARIES

To continue to improve on the summarized skimming, the team at Stony Brook created an algorithm that could summarize text. The details of training the algorithm are described in Islam et al.[6]. The team identified three key elements for the skimming algorithm; a natural language parser, a classifier, and a skimming interface.

4.1 Natural Language Parser

Using a typed dependency parse generator, called the Stanford Parser [4], sentences could be put into a tree structure. First, the words in the sentences are parsed and grammatical relations are detected between words. A parsed sentence could be shown as a set of relations holding a Governor word and a Dependent word, in the following way:

Relation (Governor → Dependent).

With the relations extracted, a directed graph structured as a tree, can be constructed. In the tree, each word is a node and the edges are the relationships between the words. The root of the tree would be a word that has no governor word and is not dependent on any other word in the sentence.

An example of one of those tree structures can be seen in Figure 1 which makes up the main structure of the figure. Section 4.2 will refer back to Figure 1 and explain more of the figure.

The Stanford Parser uses a hierarchy of 48 relations to organize the sentences into trees. In the example, the relations nominal subject and copula are higher in the tree than relations like direct object, and possession modifier, which matches the hierarchy in Marie-Catherine de Marneffe et al. [4].

4.2 Classifier

Classifiers are algorithms that predict labels using features in data (Serrano [10]). Classifying algorithms can be created through machine learning, which involves training. Classifiers can use tools like trees, graphical curves, vectors, and many others to classify data.

The researchers at Stony Brook used supervised training for the machine learning algorithm. This means that the data was labeled with classes, and the machine learning created an algorithm based on the classified data. The classifier algorithm used the training data to automatically predict

A: Gold summary with nouns only:
Twitter, 10 person startup San Francisco, Obvious. Mixture networking microblogging. idea, people omnipresence. Use Iran election.
B: Gold summary with nouns and prepositions only:
Twitter, 10 person startup San Francisco, Obvious. Mixture of networking microblogging. on idea, people omnipresence. Use in Iran election.
C: Combined gold summary:
Twitter, 10 person startup San Francisco, called Obvious. Mixture of social networking microblogging. based on idea, people enjoy virtual omnipresence. Use in Iran disputed election.
D: Original Paragraph:
Twitter, which was created by a 10 person startup in San Francisco was called Obvious. It is a heavy mixture of messaging, social networking, 'microbloging' and something called 'presence'. It's shorthand for the idea that people should enjoy an 'always on' virtual omnipresence. Twitter's rapid growth made it the object of intense interest. The object of fair amount of ridicule, as it was derided as high tech trivia of the latest in time-wasting devices. But its use in Iran in the wake of the disputed presidential election of June 2009 brought it a new respect. It was used to organize protests and disseminate information in the face of a news media crackdown.

Table 1: Example of the Gold Standard Summaries and condensed summaries [2].

Number Of Questions and Type	Question	Answer
1 question on article topic	What is the article about?	Twitter
4 questions on nouns	What is the name of the Twitter start up?	Obvious
3 questions on verbs	What was Twitter used for in Iran?	organize protests
1 question on numeric values	How many people organized Twitter?	10
1 on adjectives/adverbs	What kind of interest did Twitter generate?	Intense

Table 2: Examples of questions and answers [2].

the labels of new data. The researchers chose to label words with the classes "YES" or "NO" based on a word's appearance in a Gold Summary.

Features, or characteristics of data points, can be used to plot and organize data sets for classifying. The Stanford Parser helped the team identify features for training classifiers.

Recall Figure 1, which gives a visualization of the features of parsed sentence trees. The features the team chose for the words in the trees are part of speech, number of outgoing edges, level in the tree, number of descendants, and incoming relation type. When sentences are parsed this way the tree reveals the relative importance of words through their features.

4.2.1 Training

After selecting the features, the next step is training the classifier. The researchers used an open source machine learning library called Weka [5] and used a data set of summaries to train classifiers. The researchers created the data set with the help of 24 sighted participants. The participants summarized using guidelines similar to those in Section 3. The participants summarized 24 news articles, 674 sentences total, and summarized 3 articles each. Of the 674 summarized sentences, 591 were used for training and 83 were used for testing.

Testing means the trained algorithm is run on labeled data that the algorithm has not seen before and is evaluated on how accurate it is at classifying the data. Testing makes sure the algorithm generalizes well to previously unseen data. Testing also checks for overfitting, i.e. the issue of algorithms specializing too much to their training data.

Of the classifiers trained in Weka's library, the Support Vector Machine (SVM) classifier gave the most desirable results with a precision of 59.27%. An SVM maps out data based on features and then constructs a plane to separate data into two classes [11].

4.2.2 Refining Summaries

The previous study (Ahmed et al. [2]) demonstrated that groups of connected words aided in comprehension as opposed to disconnected words, for example "hands down" instead of "hands". Keeping relations between words intact helps with comprehension. Additionally, branches of a tree should not be disconnected from the higher levels. Adding in nodes to keep the branches connected makes the tree more cohesive.

So, after the classifier selects words for the summary, the algorithm ***MinConnectedTree*** creates a tree that has a minimum number of connections between the chosen words. The words picked by ***MinConnectedTree*** are then added to the summary.

Figure 1 has an example of ***MinConnectedTree*** adding a word to the summary text. ***MinConnectedTree*** adds the word "gained" to the summary between "businessman" and "trust" because "gained trust" is a better understood phrase than "trust" alone. Without the word "gained", "trust" would not have a connection to the rest of the tree.

With the refinements in the algorithm, the testing results for the algorithm's precision improved greatly over the results of the lone classifiers.

4.2.3 Variable Summary

For the study described in Ahmed et al. [3], the algorithm ***VariableSummary*** was created. To start, the classifier selects which words are included in the summary, and the certainty of those selections is quantified with a confidence score. A confidence score is a mathematical estimate of how certain a classifier is about the classification of a data point. The higher the score, between 0 and 1, the higher the certainty of the classifier in including a word in the summary.

Next, the words are sorted and ranked based on the confidence scores. The ranking normalizes the scores for a more consistent summarization process. Normalization being the word with the highest confidence score being given a 1.0 ranking and the word with the lowest confidence score being given a 0.0 ranking. All the other words in between would be incremented between the two.

Then, the words are arranged in the order they appear in the original sentence. An example of the ranking process can be seen in Table 3.

Finally, the words' rankings are compared to a threshold. Words that meet or exceed the threshold are included in the varied summary. Table 4 shows an example of how thresholds change the size of the summaries. The next section discusses the use of thresholds further.

4.3 User Interface

The interface created by the team at Stony Brook for touchscreen is controlled with intuitive gestures, like pinching, swiping, and dragging. The users in the study controlled the size of the summary by controlling the compression rate. The compression rate sets the threshold for the algorithm to use.

For example, if a user wanted only the most important words in the summary, they would "pinch in" to get to the 100% compression rate, setting the threshold to 1.0. If a user wanted to hear the full text, they would "pinch out" to 0%. Users preferred distinct predefined rates to switch between, rather than continuous control between 0% and 100%. The predefined rates were set to 0%, 20%, 40%, 60%, 80%, and 100%. There is audio feedback for switching between rates; different pitches sound depending on the level of compression.

Along with controlling the length of the summary, users could swipe left and right to navigate between sentences, and could drag their fingers across the screen to read the text that is directly under their finger.

In Machulla et al. [9], conveying information through a tactile medium is ranked above an audio medium but below visual mediums. For visually impaired users finding tactile alternatives to use in conjunction with audio could greatly improve the way information is conveyed. Using a touchscreen, users can directly interact with the two dimensional layout of a text and hear what is directly beneath their finger, which grants access to the specially-coded details of the text.

5. EVALUATING THE INTERFACE

The team at Stony Brook conducted a study similar to the Search Scenario described in Section 3.2 using the touchscreen interface and automated summaries. In this iteration, users skimmed/read five articles, using different methods of navigation for each. The methods were:

- Touch-based without skimming
- Touch-based with single speed (50% compression) skimming
- Touch-based with variable speed skimming
- Keyboard-based without skimming
- Keyboard-based with single speed skimming

The results, Table 5, showed that arriving at an answer with touch-based skimming was 2.3 times faster than touch-based reading (without skimming), and users answered the questions 2.2 times faster. Touch-based skimming was also faster than keyboard-based skimming.

After the Search Scenario task, the participants were given a questionnaire to be answered with a 5-point Likert scale. With 1 being "strongly disagree" and 5 being "strongly agree", the questions and average rating are:

- I wish I could look through articles faster than I can with a screen reader (4.60)
- I experience difficulties in fast navigation in an article with regular touch interface (4.13)
- Touch-based skimming made reading articles faster than regular touch navigation (4.67)
- Touch-based skimming is easier than keyboard based skimming (4.13)
- I want to use the touch-based skimming feature in the future (4.67)

The majority of participants agreed touchscreen skimming is faster and easier than both keyboard skimming and regular touchscreen navigation. The participants also wanted to use the touchscreen skimming in the future.

6. EYES REDUCED SKIMMING

Taking a different approach to skimming, team at University of British Columbia set out to address the "design problem of translating the visual interactions in skim-reading into a mode of interactions that depend less on visual attention." (Khan et al. [8])

By combining auditory and visual reading, they created an eyes-reduced design for situational impairment. Situational impairment describes the experience of reducing or losing the ability to interact with something as usual due to circumstances that occupy the user. Some circumstances that could occupy users include doing a task that takes away visual attention, like walking, not looking at a screen while commuting to avoid motion sickness, or having a migraine and being sensitive to light.

The researchers chose to focus on the context of using public transportation since commuting is growing increasingly common. With eyes-reduced design, skimming is primarily auditory but users can still look at the text as necessary.

6.1 Analysis and Design

To begin detailing design features, the researchers first conducted a study. The study focuses on situational impairment, where looking at the text would be difficult to do. Using a text to speech app called VoiceDreamReader [1], participants were given 4 texts, and were instructed to "Pretend you have a class discussion later today and you are reading the assignment on the bus but get motion sickness"(Khan et al. [8]). After the task, participants were asked comprehension questions and were interviewed.

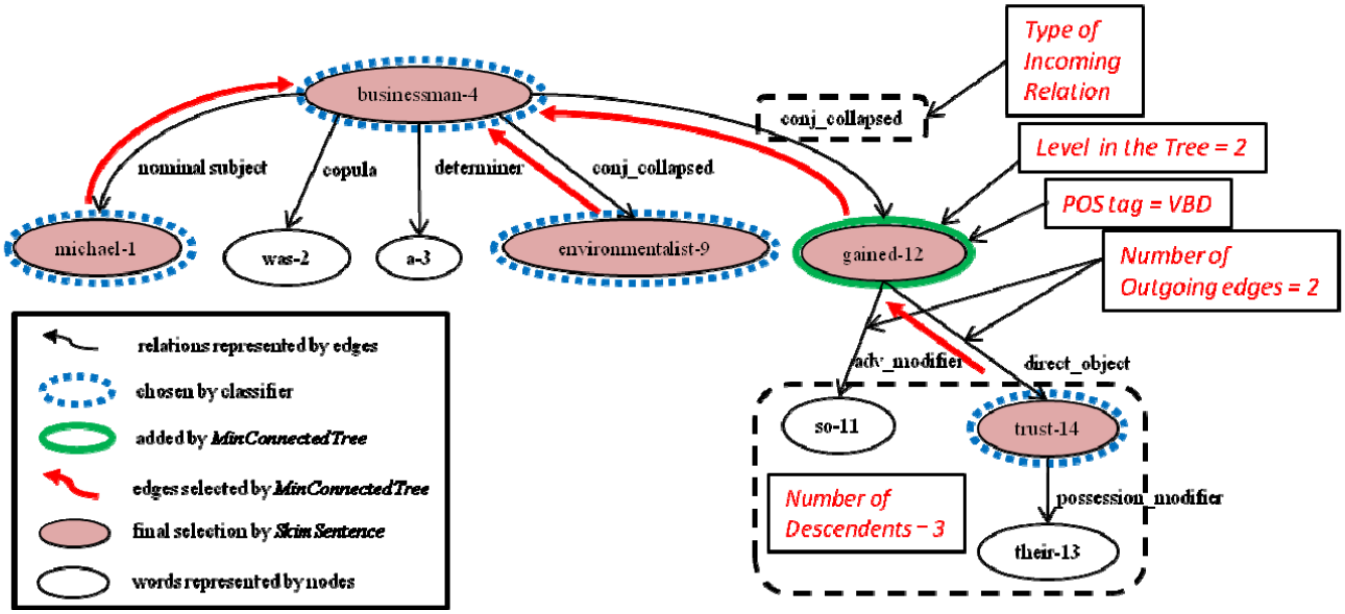

Figure 1: The sentence "Michael was a businessman, as well as an environmentalist, and so gained their trust", parsed into a tree structure, summarized. Selected features of the sentence are indicated for the word "gained". [6]

Original Sentence:	"Amy is a busy student"
Words with confidence score (SVM)	(Amy, 1.0) (is, 0.6) (a, 0.5) (busy, 0.8) (student, 0.7)
Sort and normalized score	(Amy, 1.0) (busy, 0.75) (student, 0.5) (is, 0.25) (a, 0.0)
Reorder by original position	(Amy, 1.0) (is, 0.25) (a, 0.0) (busy, 0.75) (student, 0.5)

Table 3: Example of words given ranks from confidence scores [3].

The study revealed that users wanted to jump to key sections of the text, and nonlinear navigation was difficult with VoiceDreamReader. The study also showed that the formatting and structure of a text affected comprehension.

Based on the results of the study the researchers offer eleven design guidelines for eyes-reduced skimming.

1. Provide way to navigate the structure of the article in a nonlinear fashion and to localize the current position
2. Provide semantic and spatial navigation instead of temporal navigation
3. Pause narration when the user is navigating
4. Provide ways to adjust speech rate dynamically
5. Provide ways to refer back to text content from the narration and vice versa
6. Diverge from verbatim narration for specific types of text to enhance listening comprehension e.g., break lists, announce section, expand abbreviation
7. Provide auditory or haptic feedback as nonvisual navigation cues
8. Support opt-in visual engagement
9. Support unimanual interactions (i.e. using one hand)
10. Support individual differences in skimming strategies
11. Support annotation creation and consumption

Using the design guidelines they drafted from the study, the researchers created the app *Skimmer*. For navigation, there are tabs at the very top of the screen to switch between the full text, and the Overview page. The Overview functions like a table of contents that allows users to jump to sections in the text. The Overview helps to satisfy design guidelines 1 and 5.

Following guidelines 1, 2, and 9, the touchscreen navigation works with gestures on different parts of the screen. Tapping on the left and right sides of the center of the screen navigates between sentences, swiping up and down navigates between paragraphs. Tapping the top of the screen on the left or right changes the speech rate, and tapping the bottom area navigates between discourse markers. Discourse markers are key phrases where important information is likely to be, such as "in this paper". Navigation is accompanied by audio cues, and discourse markers have a background sound to emphasize their importance. For guidelines 7 and 8, Skimmer gives haptic cues when the narration reaches a table or figure and the user can choose to look at them or continue with the narration.

6.2 Evaluation

To evaluate the design of the app, 6 graduate students were given 2 articles each to go through while riding a bus. One article was read using Skimmer, and the other using VoiceDreamReader. Once off the bus, the participants answered a questionnaire on the article and were interviewed

Summary	Threshold
Afterwards, they often spray their skin with a protective coating of dust.	0.0 (original)
Afterwards, they spray their skin a protective coating of dust.	0.2
Afterwards, they spray skin protective coating dust.	0.4
they spray skin coating dust.	0.6
spray skin coating.	0.8
spray	1.0

Table 4: How thresholds affect the size of summaries [3].

Method	Time to Reach Answer	Time to Answer
Touch-based without skimming	296.86 seconds	300.60 seconds
Touch-based with single speed skimming	148.06 seconds	150.33 seconds
Touch-based with variable speed skimming	128.07 seconds	130.93 seconds
Keyboard-based without skimming	288.27 seconds	298.80 seconds
Keyboard-based with single speed skimming	174.53 seconds	176.93 seconds

Table 5: Average times for the text to speech to read an answer out and the times the participants to answer the questions using each skimming/reading method [3].

about their experience. Comprehension between the two apps was similar, so discussion focused on the qualitative experience with the apps.

Through the interviews, it was concluded that Skimmer:

- Could be used eyes-reduced
- The Overview tab was very useful
- Auditory and haptic feedback helped users re-focus on the text
- Supported different style of navigation
- Complicated numbers and acronyms are a challenge for skimming
- Users appreciated the quality of narration and multiple voices
- Discourse markers were useful, but needing acclimation
- Figures/tables are mostly ignored, but participants appreciated the idea of a haptic nudge
- Participants appreciated Skimmer's design concept

Some of these conclusions seem to mirror design implication discussed in Machulla et al. [9], such as non-linear exploration, in the form of the Overview tab, being important to skimming. Conveying spatially coded information is again shown to be important for navigating through text.

7. CONCLUSIONS

Visually impaired users cannot skim as easily as sighted users. It is difficult to navigate through text and pick out important information with sound and touch alone.

With the Stony Brook team's work creating summaries using machine learning, important information can be extracted and manipulated at the touch of users' fingertips. Other studies built an understanding of the usefulness of navigating the structure of text, which aided in creating design guidelines for non-linear exploration through text. These studies contribute towards the goal of a generalized interface of audio accessible skimming to be used in nearly any context.

Acknowledgments

Thank you to Professor Elena Machkasova and Professor Kristin Lamberty for their support and patience throughout this project, and Charlot Shaw for the feedback.

8. REFERENCES

[1] Voice dream reader [mobile app].

[2] F. Ahmed, Y. Borodin, Y. Puzis, and I. Ramakrishnan. Why read if you can skim: towards enabling faster screen reading. *W4A '12: Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*, pages 1–10, April 2012.

[3] F. Ahmed, A. Soviak, Y. Borodin, and I. Ramakrishnan. Non-visual skimming on touch-screen devices. *IUI '13: Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 435–444, March 2013.

[4] M.-C. de Marneffe, B. MacCartney, and C. D. Manning. Generating typed dependency parses from phrase structure parses. *LREC*, 2006.

[5] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemannn, and I. Witten. The WEKA data mining software: An update, 2009.

[6] M. A. Islam, F. Ahmed, Y. Borodin, I. Ramakrishnan, T. Hedgpeth, and A. Soviak. Accessiblle skimming: Faster screen reading of web pages. *UIST*, 2012.

[7] JAWSSkimming. Skim reading, 2011.

[8] T. Khan, D. Yoon, and J. McGrenere. Designing an eye-reduced document skimming app for situational impairments. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, April 2020.

[9] T. Machulla, M. Avila, P. Wozniak, D. Montag, and A. Schmidt. Skim-reading strategies in sighted and visually-impaired individuals. *Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference*, pages 170–177, June 2018.

[10] L. Serrano. *Grokking Machine Learning.* Manning Publications Co., 2019.

[11] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer, 2000.