A Web-Based Medical Text Simplification Tool

Abstract

With the increasing demand for improved health literacy, better tools are needed to produce personalized health information efficiently that is readable and understandable by the patient. In this paper, we introduce a web-based text simplification tool that helps simplify existing text materials to make them more personalized and more broadly accessible. The tool uses features that provide concrete suggestions and all features have been shown individually to improve the understandability of the text. We provide an overview of the tool along with a quantitative analysis on medical texts. On a medical corpus, the tool provides good coverage with suggestions on over a third of the words and over a third of the sentences. These suggestions are over 40% accurate, though the accuracy varies by text source.

1. Introduction

One of the main barriers to personal health management is access to personalized health information material [1, 2, 3]. This information should be tailored to the needs of the individual both in terms of content as well as accessibility, e.g. readability. Limited understanding of health information [4] is estimated to keep 90 million Americans from obtaining health information needed to make informed health decisions [2, 5]. Limited health literacy and access to personalized information is having effect on patient behavior, ranging from changes in participation in cervical cancer screenings [6] to glycemic control in type 2 diabetes patients [7]. Without good access to health information, individual health status is reduced [5]. Financially, limited health literacy can cause patients to delay preventative care and make inappropriate decisions and is estimated to cost up to $238 billion annually [8].

Text remains one of the most cost-effective and informative ways of distributing information to patients. It can been used in a wide range of settings (pamphlets, e-mails, webpages, mobile devices, etc.) and has be used successfully to inform patients on a variety of types of topics including to teen alcohol use [9], appropriate sunscreen usage [10], and cardiovascular health [11]. For these reasons, in this work we focus on text as a medium for information dissemination.

There are two main challenges for generating personalized health information texts. First, generating the text is cost-prohibitive and requires a medical professional or medical writer time to generate. Second, the text must be understandable by the text consumer, which often have different background and education that text producers. While medical professionals are given some high-level guidance on how to simplify text (e.g., encouraged to use "plain language" [12, 13]), intuition is not always enough to produce texts that are understandable for a particular patient.

Unfortunately, few tools exist to solve these two problems. Historically, the most common "tools" that have been provided are readability formulas, for example, the Flesh-Kincaid readability formula (and similar variants) [14] or the Simple Measure of Gobbledygook [15]. These formulas provide a single numerical score for a text, often based on fairly rudimentary statistics, e.g., average number of syllables per word and average sentence length. The readability formulas have mostly been validated based on corpus analyses, i.e. showing that texts already written at known readability levels tend to correlate well with the readability scores. However, no research has shown that simplifying text guided by readability formulas produces text that is easier to understand for the end user and some studies have shown that simplifying texts based on these formulas is problematic [16], particularly in the medical domain [17].

In this paper, we introduce a medical text simplification tool that provides concrete guidance on what text portions are problematic and how to fix them. The tool is freely available and helps create personalized health text more efficiently that is of higher quality. We outline the design goals used in developing the tool and then examine the individual feature components used within the tool to guide simplification. Critically, all features used in the tool have been individually validated in previous experiments. Only those features that are effective at improving readability and understanding are included in the tool. While all features have been tested individually, they have not been tested in aggregate. We conclude with an analysis of the overall coverage of the tool and the quality of the suggestions made.

2. Design Goals

The development of the text simplification tool has been guided by a number of overarching design goals. First, we wanted to tool to have be broadly accessible with minimal overhead for use. The tool should be able to be used by users in different settings on different operating systems with minimal setup requirements. For this reason, we chose to make the tool available online as a web resource. The tool does not require any additional software (e.g., as a Microsoft Word plug-in would) and only requires a modern browser. Like most modern web applications, a majority of the computation, etc. is done on the server side, so the hardware requirements are also minimal.

Second, the tool should be stand-alone and the user should be able to get guidance and edit the document interactively. Specifically, we integrated all of the text simplification features into an interactive text editor. This allows the user to get feedback from the tool and edit the text at the same time. Third, any features that are used in the tool should be experimentally validated. The features we use have been shown not only to have an impact on text readability, but to help produce text that is perceived as easier to readers and that improves understanding.

To validate our features, we often take a multistep approach. Potential features are first identified and validated by corpus analyses. For some features, an initial experimental validation is done focusing on the feature in isolation. For example, using word frequency as a proxy for word difficulty has been validated by comparing the proportion of users that know the definition of a word with the frequency of that word' in a large corpus [18]. Finally, the feature is then evaluated when used in a realistic setting where it is used to simplify actual text. The text simplified with the feature is then compared against the original text in a user study evaluating perceived difficulty, understandability, and recall. Only those features that pass this rigorous experimentation and show positive impact should be integrated into the tool.

Fourth, the tool should provide concrete advice to the user about what parts of the text are problematic *and* how to improve these areas. As a counterexample, readability formulas generally do not provide concrete guidance on what parts of the text should be changed and, critically, what they should be changed to.

Fifth, once the user has finished simplifying the document, they should understand the impact of the changes that they have made. In addition to the editor, we also provide a post-simplification analysis of the text and show how it differs from the original text along a number of dimensions that have been shown to impact understanding. This report helps the user see where changes were made and how they affected the text.

Finally, this tool is a research tool. As such, it is critical that the tool can be properly evaluated and used to motivate future research. To facilitate this, the tool was designed such that as much data could be collected about how the tool was used. This data will be used to understand how the tool is used and provide data to further refine the tool, e.g., by using machine learning.

3. Text Simplification Features

The most important components of the tool are the text simplification features that are used to identify problematic text and make suggestions for improvement. The features use in the tool can be broken down into two general categories: word-level features and sentence-level features.

3.1. Word-level features

At the word level, the tool identifies difficult words and then provides either candidate replacements or guidance on how that word could be replaced. The tool only provides suggestions for difficult, content-bearing words. We identify content-bearing words based on their part of speech as identified by an automatic parser [19], specifically, nouns, verbs, adjectives and adverbs. Difficult words are identified based on their frequency in the Google Web Corpus [20]. These are words that may not be understood by all readers and are candidates for changing. Previous work has shown that frequency in a large corpus is a good measure of word difficulty [21, 18]. By default, words are identified as difficult if they are not in the 5,000 most frequent words in the corpus. Different target audiences will have different backgrounds and different vocabularies, so the tool also contains a slider (see Figure 1) to adjust the difficulty level threshold to identify less words as difficult (increasing the threshold), or, more often, identifying more words as difficult (decreasing the frequency threshold).

For each word that is flagged as difficult, we look up information that we have about that word that can help the user simplify the word. Information about words has two general categories, direct substitutions and word change guidance. Direct substitutions are resources where we can directly suggest candidate word substitutions to replace the difficult word (e.g., synonyms). We use three sources for direct substitution.

WordNet

WordNet is a database of nouns, verbs, adjectives and adverbs [22]. For a given word and part of speech, the word is further subdivided based on possible senses, i.e. meanings, of the word. For example, the word "heart" has 10 different senses for the noun, with the first two senses being 1) the source of feelings as in "follow your heart" and 2) the physical organ. These word senses are then grouped into synonym groups. We use these synonym groups to identify alternate substitutions for a difficult word. Specifically, given a difficult word, we lookup all possible word senses of the word that match the part-of-speech of the word based on an automatic parsing of the sentence. Each word sense denotes a possible interpretation of the word. For each word sense, we get possible synonyms of the word. For example, the possible synonyms for heard (the physical organ) include "pump" and "ticker".

We aggregate all of the possible synonyms from all the senses of the word for the given part of speech. These represent possible substitutions for the original word. Finally, we only want to suggest synonyms that are simpler than the original word. We again use frequency to measure this and filter the list of synonyms to include only those that have *higher* frequency than the original word, representing simpler substitutions.

UMLS Metathaurus

Like WordNet, the UMLS Metathaurus is a collection of words organized by concepts (similar to senses) and their relationships [23]. Each word may have one or more concepts that it is associated with. Importantly, the Metataurus contains words that are biomedical and is therefore particularly suited for the target medical audience of the tool. Given a difficult word, we lookup all concepts related to that word. For each concept, we can then do the reverse and find all words that are related to that concept. We aggregate all of the words for each concept associated with the difficult word as candidate synonyms and again filter them to only include words that are easier (higher frequency) than the original difficult word.

Negation

In English, some word prefixes can indicate negation of the rest of the word. For example, the prefixes "ab-" and "un-" can indicate negation, i.e., *abnormal* means "not normal" and *unbalanced* means "not balanced". Additionally, some suffixes can also indicate negation, for example, the suffix "-less"; the word *endless* means "without end". We use the negation parser from [24] to identify words that are morphological negations and can be broken down into either "not" or "without" plus another word. The negation parser uses a large lexicon of words along with hand-crafted rules to identify those words. If a difficult word is identified that is negative, we include as a simplification candidate the positive version of the word with the appropriate simple negation word (e.g., "not balanced" for *unbalanced*). Often, the positive version of the word is more accessible (e.g., *normal* vs. *abnormal*) and may be easier to understand by the reader.

These three resources provide direct suggestions to the user that can be used to replace a difficult word in the context of the sentence. We also included two other sources of information that do not provide direct substitution options, but give guidance to the user to help them simplify the word and/or the sentence the word occurs in.

Nominalization

Nominalization is when a word that has a non-noun form is used as a noun, for example, *misadjustment* is the noun form of the verb *misadjust*. Previous work has shown that difficult text contains more nouns, while easy text contains more verbs [25]. Additionally, a common guidance when simplifying text is to use the active voice, rather than the passive. Motivated by these, we identify nominals in the text and suggest they be rewritten as the corresponding verb. We use the nominalization database of the UMLS [26]. For any noun in the database where there is a corresponding verb, we add possible guidance for the user to consider using the verb form rather than the noun form.

Affix analysis

The negation parser is an example of using the morphology of the word to create a simpler variant. This idea can be generalized beyond just negation to break a word into its morphological parts (called affixes), i.e., prefixes, stems, and suffixes. The stem of the word is the main content bearing part of the word. Prefixes and suffixes modify and/or elaborate on the meaning of the stem. For example, *cardiovascular* can be broken down into the prefix *cardio*, referencing the heart, and the stem *vascular*, referencing blood vessels. For difficult words where we do not have information about the original word, but do have information about the affixes, we can provide some information about the word. Using the affixes, we generate an explanation for the meaning of the word automatically that can help simplify the word. For example, for *cardivascular*, we generate the explanation "relating to the heart and blood vessels" based on *cardio* and *vascular*.

We use the SubSimplify algorithm [27] to generate possible explanations for difficult words based on morphological analysis. The tool uses a lexicon of 586 affixes with known definitions combined with stem definitions from WordNet [22] and the Consumer Health Vocabulary [28] to generate word explanations. Previous work showed that while the coverage of the SubSimplify algorithm was good, the quality of the suggestions was not as good as other resources, i.e., WordNet and the UMLS Metathaurus. Because of the quality and because this analysis only provides explanations and not actual substitutions, we only show affix explanations if none of the other features have suggestions.

3.2. Sentence-level features

In addition to the word-level features, we also identify sentence-level characteristics that could be changed to improve the understandability/readability of the text.

Double negatives

A double negative occurs in a sentence when you have two negated terms in reference to the same concept. For example, the sentence

It is not illogical to carry out the study.

contains two negations, *not* and *illogical* (e.g., not logical). Often sentences with double negatives can be written more clearly by rewriting the sentence in the positive, e.g.,

It is logical to carry out the study.

There are two main sources of double negatives in text. The most common is when a negative word like *no*, *nor*, or *not* is combined with a word that is morphologically negative (e.g., illogicial). Alternatively, some sentences can contain two negatives, e.g.,

The hospital won't allow no more than one visitor.

This is less common, particularly in well-written text.

We again use the negation parser from [24] to identify double negatives. The tool combines morphological negation identification (as described previously) with a parse tree based approach to identify sentences with double negation.

Grammar simplification

One common suggestion for simplifying text is to use "simpler" grammatical structures. Unfortunately, most tools and readability formulas give little or no guidance on exactly how to accomplish this beyond looking at sentence length. To assist the editor in simplifying sentence grammar, we integrated the grammar rules from [29].

The grammar rules consist of two parts: 1) a difficult parse subtree that may be problematic (left hand side) and 2) a corresponding parse subtree that the difficult subtree should be transformed into (right hand side). For example, the rule:

$$(NP(DT)) \rightarrow (NP(DT)(NN))$$

applies to a noun phrase that is only a determiner in a difficult sentences and suggests simplifying the sentence by including an additional noun in the noun phrase to make the context more clear. For example, the sentence:

This was added on to treatment with beta blockers.

has the following parse tree (indentation has been used to indicate levels in the parse tree):

```
(S
  (NP (DT This))
  (VP (VBD was)
    (VP (VBN added)
      (PP (IN on)
        (PP (TO to)
          (NP (NN treatment))))
      (PP (IN with)
        (NP (JJ beta) (NNS blockers)))))
  (. .))
```

This parse tree does have a noun phrase that is only a determiner, so the grammar rule would apply. The user is provided guidance in the form of a description of what the rule transformation suggests (in a non-technical form) as well as an example transformation. For example, for the grammar rule above, the user would be given the rule explanation:

```
Make sure the pronoun is clear or
defined in the previous sentence.
```

along with an example:

This was added to care instructions later.

This dosage was added to care instructions later.

Given this guidance, an example simplification of the sentence above might be:

This therapy was added on to treatment with beta blockers.

We included 146 rules and identified applications of the rule based on the parse string of the each sentence in the input text.

4. Softward Details

We chose to make the tool web-based to maximize accessibility and minimize barriers to entry. When developing the tool, we reused as many existing tools/frameworks as possible to minimize the amount of custom code written. Additionally, it is important that we can eventually distribute the code and that other researchers can replicate the tool without a huge overhead, so we only used resources that were open source or freely available.

For the backend, we use the Spring web framework[1], which is a Java-based application framework. When using it as a web framework, the framework generates a web server based on Apache Tomcat. We use a Mysql database to store both resources (e.g., Google web frequencies, UMLS resources, etc.) as well as to store usage data. We use the Stanford CoreNLP toolkit [19] for all text processing, including preprocessing (sentence splitting and tokenization), syntactic parsing, and dependency parsing.

The frontend is written in Javascript with html/css for content and formatting. We use the Codemirror text editor [2] for editing and highlighting the text. It provides a reasonable in-browser text editor with standard functionalities and allows for customizable highlighting/formatting. Custom Codemirror modes were written to support word-level and sentence-level highlighting based on the suggestions from the backend.

5. Tool Walkthrough

The current working version of the tool can be found online[3]. Figure 1 shows a screenshot of the tool during editing. The tool consists of three main sections. The majority of the tool is taken up by the text editor, which supports free editing of the text and standard editor operations (copy, paste, undo, etc.). On the right side of the tool are stats about the text (top portion) and editor configuration (bottom portion). The bottom of the tool (blue bar) contains sentence-level guidance when the user clicks on a sentence that could be improved. Finally, at the very bottom of the tool are three buttons.

When the user first visits the webpage the editor is empty and there is no sentence-level guidance. The user then types or, more likely, pastes the text to be simplified into the editor and clicks “Simplify”. This sends the text to the backend for processing. The backend processes the text, applies all the word-level and sentence-level features to the text, and sends the relevant information back to the frontend. All difficult words where candidate simplifications were generated by the tool are underlined and showed in blue. All sentences/sentence portions where sentence simplification guidance is possible are underlined and italicized.

Figure 1 shows the tool at this stage in the process. For this text, 38 words have been identified as difficult where the system has suggestions. The user can click on any of these highlighted words and a dropdown menu appears with the possible options. Options are color coded according to their source to give the user more information about the usefulness of individual features. For example, Figure 2 shows the menu that would be shown if the user clicked on “bronchoconstriction”. The first option is from the UMLS and the second option from nominals. The user can click on any of the options to have the word replaced with that option (just the verb in the case of nominals). After clicking, the user can still edit the text normally to fix the word or the context to make the suggestion more appropriate.

Figure 1 also shows four sentences in the text that have guidance based on the grammar rules (“difficult for air”, “otherwise healthy”, “This”, and “allergies or asthma”). If the user clicks on any of these highlighted text portions the associated rule is populated at the bottom of the tool. The rule is shown to the user in English and an example simplification (both the original and simplification) are shown with the changed parts highlighted. In this case, “otherwise healthy” was clicked on and the rule suggestion is shown (suggesting, e.g., changing this to something like “healthy”).

5.1. Statistics

The user can continue to edit their document based on the tool feedback until they are satisfied with the document. Once the user is done, the tool also provides summary statistics to give the tool user some feedback on how the changes that they made affected the text. When the user clicks on the “Get Stats” button, a separate tab opens up in the browser that shows summary statistics for the original document alongside the finalized version of the document. Figure 3

[1] https://spring.io/
[2] https://codemirror.net/
[3] anonymized

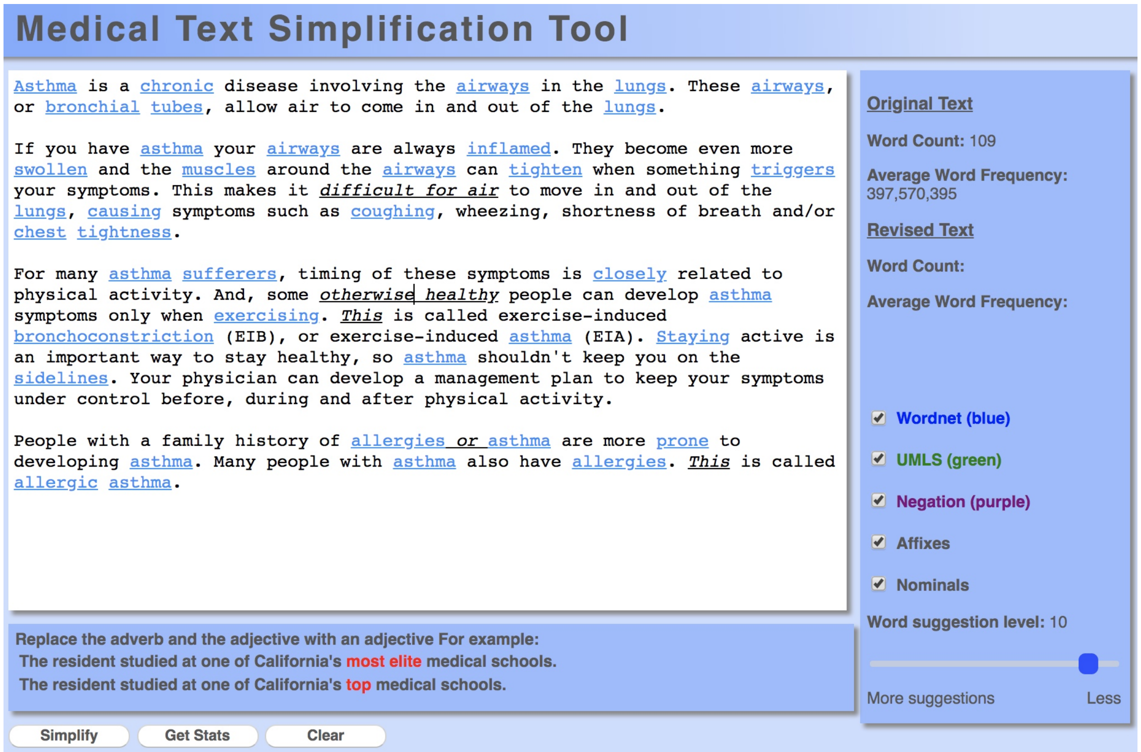


Figure 1. A screenshot of the tool after entering text and clicking the "Simplify" button. The underlined text "otherwise healthy" has been clicked causing the associated grammar rule to show up at the bottom.

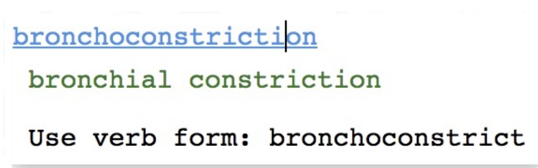


Figure 2. Example dropdown menu for word-level simplification.

shows a screenshot of an example of this. The statistics table contains a comparison of the original versus the simplified for: text, number of sentences, average sentence length, average word frequency (of content-bearing words), number of words, number of nouns, and number of verbs.

5.2. Customization

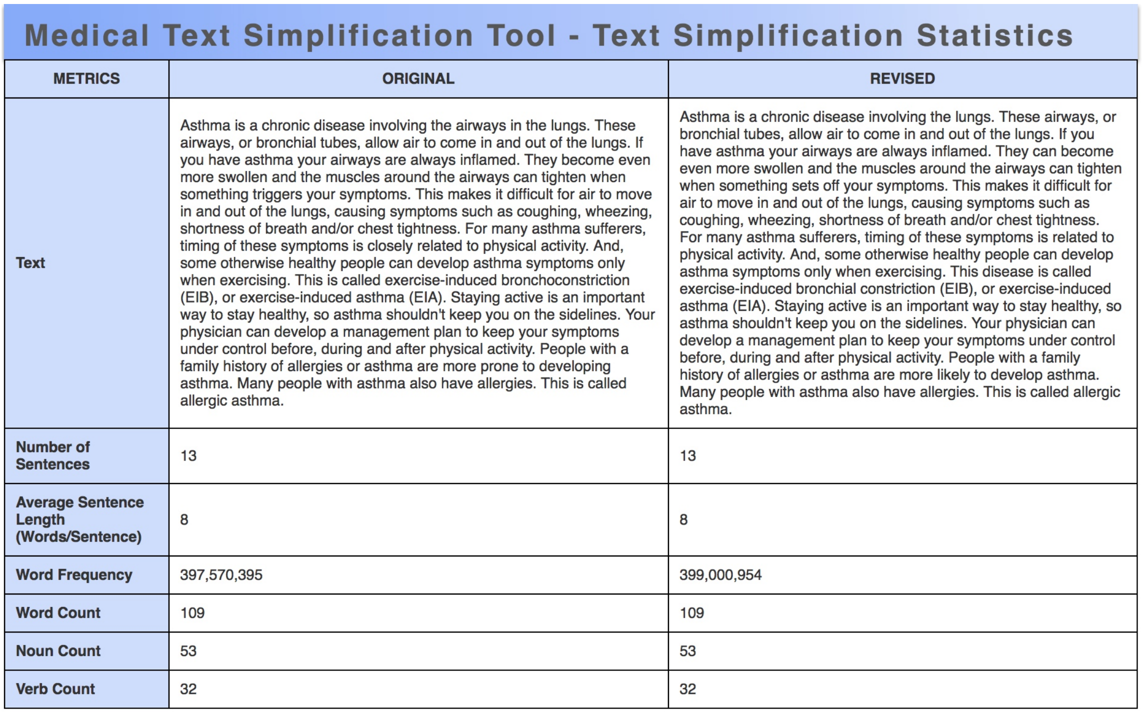
The tool allows for some customization of the simplification features applied (bottom right of the tool). Each of the five word-level features are by default applied to the text, but the user can deselect any of the tools to omit the suggestions. The other main customization that the user has is to adjust the level for which words are considered difficult. By default, the editor identifies any word that does not occur in the 5,000 most frequent words from the Google web corpus as difficult. The slider can be adjusted right to flag fewer words as difficult (corresponding to words not in the 10,000 most frequent words). Or, more likely, the user can choose to get suggestions for more words by moving the slider to the left. Each move left on the slider corresponds to adjusting the ranking by 500, i.e. one move to the left flags any word as difficult that isn't in the 4,500 most frequent words. two moves 4,000 most frequent, etc.

6. Tool Analysis

To help understand the impact the tool will have for simplifying medical text we conducted experiments to measure how frequently suggestions are made and the quality of those suggestions.

6.1. Tool coverage

To measure how often simplification guidance was available, we applied the tool to a medical corpus. We collected 195 medically related articles from English Wikipedia. We chose Wikipedia since this is a common source for information online and is frequently returned



Medical Text Simplification Tool - Text Simplification Statistics

METRICS	ORIGINAL	REVISED
Text	Asthma is a chronic disease involving the airways in the lungs. These airways, or bronchial tubes, allow air to come in and out of the lungs. If you have asthma your airways are always inflamed. They become even more swollen and the muscles around the airways can tighten when something triggers your symptoms. This makes it difficult for air to move in and out of the lungs, causing symptoms such as coughing, wheezing, shortness of breath and/or chest tightness. For many asthma sufferers, timing of these symptoms is closely related to physical activity. And, some otherwise healthy people can develop asthma symptoms only when exercising. This is called exercise-induced bronchoconstriction (EIB), or exercise-induced asthma (EIA). Staying active is an important way to stay healthy, so asthma shouldn't keep you on the sidelines. Your physician can develop a management plan to keep your symptoms under control before, during and after physical activity. People with a family history of allergies or asthma are more prone to developing asthma. Many people with asthma also have allergies. This is called allergic asthma.	Asthma is a chronic disease involving the lungs. These airways, or bronchial tubes, allow air to come in and out of the lungs. If you have asthma your airways are always inflamed. They can become even more swollen and the muscles around the airways can tighten when something sets off your symptoms. This makes it difficult for air to move in and out of the lungs, causing symptoms such as coughing, wheezing, shortness of breath and/or chest tightness. For many asthma sufferers, timing of these symptoms is related to physical activity. And, some otherwise healthy people can develop asthma symptoms only when exercising. This disease is called exercise-induced bronchial constriction (EIB), or exercise-induced asthma (EIA). Staying active is an important way to stay healthy, so asthma shouldn't keep you on the sidelines. Your physician can develop a management plan to keep your symptoms under control before, during and after physical activity. People with a family history of allergies or asthma are more likely to develop asthma. Many people with asthma also have allergies. This is called allergic asthma.
Number of Sentences	13	13
Average Sentence Length (Words/Sentence)	8	8
Word Frequency	397,570,395	399,000,954
Word Count	109	109
Noun Count	53	53
Verb Count	32	32

Figure 3. A screenshot of the tool after entering text and clicking the "Simplify" button. The underlined text "otherwise healthy" has been clicked causing the associated grammar rule to show up at the bottom.

Documents	195
Sentences	11,568
Content words	216,081
Words with simplifications	74,480 (34%)
from WordNet	34,946 (16%)
from UMLS	29,838 (14%)
from Negation	1,623 (0.75%)
from Affix	21,775 (10%)
from Nominalizations	4,944 (2.3%)
Double negative sentences	150 (1.3%)
Grammar rule sentences	4,361 (38%)

Table 1. Statistics from the medical Wikipedia corpus for how frequently word and sentence level suggestions were available.

in the top search results; The corpus contains 11.5K sentences and 216K words. To measure tool coverage, we calculated how frequently word simplifications were available (both in aggregate, as the tool is generally used, and per feature) and how frequently sentence guidance was available. We used the default threshold to identify difficult words, i.e., words that are not in the top 5,000 most difficult words. We only report statistics on content-bearing words (nouns, verbs, adjectives, and adverbs).

Table 1 shows the results of the study. Overall, the tool was able to provide suggestions on many of the words and sentences in the corpus. Over a third of the content-bearing words had suggestions from the tool. Wikipedia and the UMLS were the most productive resources (16% and 14% of the words had candidate simplifications, respectively) followed by Affix analysis (10%). Nominalization and negations had much smaller impacts. Affix explanations were only offered when no other options existed from the other sources, so 24% of the words had coverage from the other four sources. WordNet and UMLS offered some candidates on the same word, but were also complementary making suggestions on a large number of different words.

At the sentence level, the grammar rules provide good coverage with over a third of the sentences having suggestions based on the rules. Double negations were more rare and occurred in just over one percent of the sentences.

6.2. Word substitution quality

In addition to tool coverage, we also evaluated the quality of the tool suggestions. One of the most important components of the tool is the word suggestion since this provides not just guidance, but candidate replacements, and improving vocabulary is a common operation when simplifying text. We manually examined the word suggestions made for medically related documents from English Wikipedia and summaries from Cochrane[4]. We chose these two document sources since they are common resources used online for patients. We selected seven documents from each source and then evaluated how UMLS and WordNet performed, the two features that provide simpler synonyms for difficult words.

Table 2 shows the number of suggestions made by UMLS and WordNet across the two sources as well as the appropriateness. A candidate substitution was judged as appropriate if it had roughly the same meaning as the difficult word in the context of the sentence. Like the tool coverage study, WordNet also had better coverage than the UMLS on these 14 document with suggestions for almost twice as many words. The quality of the suggestions, however, was fairly comparable over the two sources with 43% of the suggestions being appropriate for WordNet and 39% for UMLS.

The source of the text affected the quality of the suggestions with Cochrane articles resulting in better options, particularly for UMLS. UMLS suggestions were appropriate 70% of the time for Cochrane articles (vs. 30% for Wikipedia) and WordNet suggestions were appropriate 55% of the time for Cochrane (vs. 33% for Wikipedia).

7. Future Work

There are many areas to explore based on the current state of the tool. Most importantly, the tool needs to be better evaluated to see the impact in realistic use cases. All tool features have been individually evaluated in independent experiments. However, this paper presents the first results of the features as integrated into the tool. Further experimentation is required to measure the individual impact of the features in the tool both intrinsically, e.g. as done above measuring coverage and quality, but more importantly, extrinsically, measuring the impact of the tool on text simplification. We are currently planning two studies to examine the effectiveness of the tool on the tool users (e.g., medical writers) as well as consumers of the content generated by the tool (patients). Based on the results of these studies we will release an updated version of the tool and advertise the tool more broadly.

Ideally, the tool will begin to be used to help guide text simplification. As noted, we are collecting usage information in a database. This data should prove invaluable for helping future tool development. It will help understand how the tool is used, what features are the most effective, and what types of texts users simplify. Additionally, the data can be used to train machine learning models to better refine the tool, e.g., candidate option ranking.

There are still additional features that can be integrated into the tool to further improve the type of guidance and direction the tool gives. The current tool gives word-level and sentence-level guidance, but document-level guidance that helps direct flow between sentences should be integrated. For example, transition words (“however”, “for example”, “first”) are critical to the flow of text and indicate how sentences relate. We hope to support transition word suggestion to indicate where transition words could help improve the flow of the text.

[4]https://www.cochrane.org/

References

[1] D. G. Simons-Morton, P. D. Mullen, D. A. Mains, E. R. Tabak, and L. W. Green, “Characteristics of controlled studies of patient education and counseling for preventive health behaviors,” *Patient Education and Counseling*, 1992.

[2] D. A. Kindig, A. M. Panzer, L. Nielsen-Bohlman, *et al.*, *Health literacy: a prescription to end confusion*. National Academies Press, 2004.

[3] K. Eichler, S. Wieser, and U. Brügger, “The costs of limited health literacy: a systematic review,” *International journal of public health*, 2009.

[4] B. D. Weiss, *Health literacy and patient safety: Help patients understand. Manual for clinicians*. American Medical Association Foundation, 2007.

[5] X. Yan, D. Song, and X. Li, “Concept-based document readability in domain specific information retrieval,” in *Proceedings of Information and Knowledge Management*, 2006.

[6] S. Garbers and M. Chiasson, “Inadequate functional health literacy in spanish as a barrier to cervical cancer screening among immigrant latinas in New York City,” *Preventing Chronic Disease*, 2004.

[7] C. Powell, E. Hill, and D. Clancy, “The relationship between health literacy and diabetes knowledge and readiness to take health actions,” *The Diabetes Educator*, 2007.

[8] J. A. Vernon, A. Trujillo, S. Rosenbaum, and B. DeBuono, “Low health literacy: Implications for national health policy.” `http://www.pfizerhealthliteracy.com/asset/pdf/Low-Health-Literacy_Implications-for-National-Health-Policy.pdf`, 2007.

[9] S. C. Banerjee, K. Greene, M. L. Hecht, K. Magsamen-Conrad, and E. Elek, “?drinking won’t get you thinking?: a content analysis of adolescent-created print alcohol counter-advertisements,” *Health communication*, 2013.

[10] B. J. Walkosz, D. B. Buller, P. A. Andersen, M. D. Scott, M. B. Dignan, G. R. Cutter, X. Liu, and J. A. Maloy, “Dissemination of go sun smart in outdoor

	Wikipedia		Cochrane		Combined	
	WordNet	UMLS	WordNet	UMLS	WordNet	UMLS
Words with suggestions	123	66	105	54	228	120
Appropriate suggestions	40 (33%)	20 (30%)	58 (55%)	27 (70%)	98 (43%)	47 (39%)

Table 2. Number of suggestions that are appropriate for WordNet and UMLS on a corpus of 14 document, 7 from Wikipedia and 7 from Cochrane.

recreation: effect of program exposure on sun protection of guests at high-altitude ski areas," *Journal of health communication*, 2014.

[11] A. G. Brega, K. A. Pratte, L. Jiang, C. M. Mitchell, S. A. Stotz, C. LoudHawk-Hedgepeth, B. D. Morse, T. Noe, K. R. Moore, and J. Beals, "Impact of targeted health promotion on cardiovascular knowledge among american indians and alaska natives," *Health education research*, 2013.

[12] A. D. Adkins and N. N. Singh, "Reading level and readability of patient education materials in mental health," *Journal of Child and Family Studies*, 2001.

[13] H. M. Brandt, D. H. McCree, L. L. Lindley, P. A. Sharpe, and B. E. Hutto, "An evaluation of printed hpv educational materials," *Cancer Control*, 2005.

[14] J. Kincaid, R. Fishburne, R. Rogers, and B. Chissom, "Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for navy enlisted personnel," *Research Branch Report 8-75. Chief of Naval Technical Training: Naval Air Station Memphis*, 1975.

[15] K. L. Grabeel, J. Russomanno, S. Oelschlegel, E. Tester, and R. E. Heidel, "Computerized versus hand-scored health literacy tools: a comparison of simple measure of gobbledygook (smog) and flesch-kincaid in printed patient education materials," *Journal of the Medical Library Association: JMLA*, 2018.

[16] C. Zarcadoolas, "The simplicity complex: exploring simplified health messages in a complex world," *Health promotion international*, 2010.

[17] M. A. Bailey, P. A. Coughlin, S. Sohrabi, K. J. Griffin, S. T. Rashid, M. A. Troxler, and D. J. A. Scott, "Quality and readability of online patient information for abdominal aortic aneurysms," *Journal of vascular surgery*, 2012.

[18] G. Leroy and D. Kauchak, "The effect of word familiarity on actual and perceived text difficulty," *Journal of American Medical Informatics Association*, 2013.

[19] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit.," in *ACL (System Demonstrations)*, 2014.

[20] T. Brants and A. Franz, "Web 1T 5-gram version 1." Linguistic Data Consortium, Philadelphia, 2006.

[21] G. Leroy, J. E. Endicott, D. Kauchak, O. Mouradi, and M. Just, "User evaluation of the effects of a text simplification algorithm using term familiarity on perception, understanding, learning, and information retention," *Journal of Medical Internet Research (JMIR)*, 2013.

[22] G. A. Miller, "Wordnet: A lexical database for english," *Communications of the ACM*, 1995.

[23] A. R. Aronson, "Effective mapping of biomedical text to the umls metathesaurus: the metamap program.," in *Proceedings of the AMIA Symposium*, p. 17, 2001.

[24] P. Mukherjee, G. Leroy, D. Kauchak, S. Rajanarayanan, D. Y. R. Diaz, N. P. Yuan, T. G. Pritchard, and S. Colina, "Negait: A new parser for medical text simplification using morphological, sentential and double negation," *Journal of biomedical informatics*, 2017.

[25] G. Leroy and J. E. Endicott, "Combining nlp with evidence-based methods to find text metrics related to perceived and actual text difficulty," in *Proceedings of t SIGHIT International Health Informatics Symposium*, 2012.

[26] S. B. Johnson, A. Aguirre, P. Peng, and J. Cimino, "Interpreting natural language queries using the umls.," in *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 1993.

[27] N. Kloehn, G. Leroy, D. Kauchak, Y. Gu, S. Colina, N. P. Yuan, and D. Revere, "Improving consumer understanding of medical text: Development and validation of a new subsimplify algorithm to automatically generate term explanations in english and spanish," *Journal of medical Internet research*, 2018.

[28] Q. T. Zeng, T. Tse, G. Divita, A. Keselman, J. Crowell, and A. C. Browne, "Exploring lexical forms: first-generation consumer health vocabularies," in *AMIA Annual Symposium Proceedings*, 2006.

[29] A. Szep, M. Szep, G. Leroy, D. Kauchak, N. Kloehn, D. Revere, and M. Just, "Algorithmic generation of grammar simplification rules using large corpora," in *Proceedings of AMIA Informatics Summit*, 2019.