Demo: An Online Evidence-based Text Simplification Editor for Medical Text

David Kauchak[1], Gondy Leroy[2], and Melissa Grueter[1]

[1]Computer Science, Pomona College, [2]Management Information Systems, University of Arizona

Abstract

It has become increasingly common for patients to turn to online resources to find medical and healthcare related information. This reflects recent advances in medicine as well as the large amount of educational resources available. Since many consumers use the Internet for reading medical information, providing this information optimized for non-specialists is essential. We are developing a medical text editor for use by medical professionals to support writing text optimized for comprehension and retention by average readers. We have developed several data-driven algorithms that identify difficult text and make suggested improvements. We evaluated these for their effect on text difficulty and on improving reader comprehension and have integrated those algorithms that were successful into an online prototype. To our knowledge, we are the first to provide an evidence-based text editor that combines algorithms shown to affect text difficulty and the resulting comprehension and retention.

Introduction

Today, we see a consistent rise in the number of patients with chronic diseases, which is requiring more and more patients to manage their own health and condition. New developments in biomedicine and precision medicine are leading to new treatments and cures that often require more patient involvement and better patient education. Additionally, many health consumers are self-educating and using the Internet to be informed and educated about healthy lifestyles. These groups benefit from increasing their knowledge of their condition and healthcare, i.e., increasing their health literacy. These all are leading to an increasing need for better health information and more effective ways of educating patients through educational material. Improving health literacy has been argued to be essential for the Patient Protection and Affordable Care Act to be successful and has been identified by the Healthy People 2020 statement by the Department of Health and Human Services as an important national goal (HC/HIT-1).

Our goal is to create tools that help optimize text for patient education since text is relatively easy to create, compared to other media such as video or interactive tutorials, and since 80% of online users [1] from different backgrounds [2] read health-related text on the Internet. We are developing an online medical text editor that combines several algorithms that have been experimentally shown to create text that is easier to understand and remember. Currently, very few tools exist to generate optimized medical educational materials for the general public and those that are available provide little concrete guidance (often only indicating the difficulty level of the document), miss evaluations of effectiveness, and only address a limited set of text characteristics. For example, readability formulas, such as the Flesch-Kincaid formula, were used early on, especially in the first decade of 2000, but there is little evidence that such formulaic outcomes can be used to measure comprehension and retention. Other tools require manual evaluation of a text, e.g., RAIN [3], or indicate difficulty but without suggestion alternatives, e.g. work that estimates text coherence [4] or the Coh-Metrix tool [5] . Our editor highlights difficult sections in text at the term, phrase and sentence level, and provides concrete alternatives and guidance on how to simplify that section using completely automated algorithms. The writer may then choose to apply or ignore the tool suggestions.

We will demonstrate the first prototype of the editor which is fully functional and available online. We have already integrated those algorithms that have shown to be most effective and will continue to augment our editor with additional algorithms. We are conducting usability studies for the tool interface with both novice and expert medical writers and will be updating the interface based on the results. Our demonstration will focus on our English language editor, however, several of our algorithms have been developed and tested for both English and Spanish and Spanish will supported in the next version.

Prototype

Development Life Cycle

Figure 1 provides a high-level overview of the editor creation development process. The prototype combines different algorithms that each have been developed and validated independently. We first identified candidate text simplification algorithms motivated by suggested common best practices, a thorough literature review, and corpus statistics studies comparing easy and difficult text. We further restricted candidate algorithms to those that could be fully automated. For each candidate algorithm, we tested their effectiveness by applying them to medical texts and evaluating their individual impact on comprehension and retention of information through user studies. Only algorithms that positively affected comprehension and retention were integrated into the editor.

Once we had a collection of different algorithms that affect text difficulty individually, we then explored how best to integrate all of them into a single working tool. Algorithms were roughly grouped by functionality into word-level components and sentence-level components and then integrated into a shared online interface. Some algorithms have been fully integrated and validated, while others have only partial functionality. Figure 1 shows a blue bar indicating integration progress for each of the algorithm components. For example, the term familiarity algorithm has been developed and tested with user studies and has been optimized for integration in the editor. It is close to the end of its development cycle with the exception of occasional updates of the underlying lexicons. In contrast, the lexical chain algorithms has been developed and shown to be effective in identifying complex sections in a text, but integration has only been completed on the editor backend.

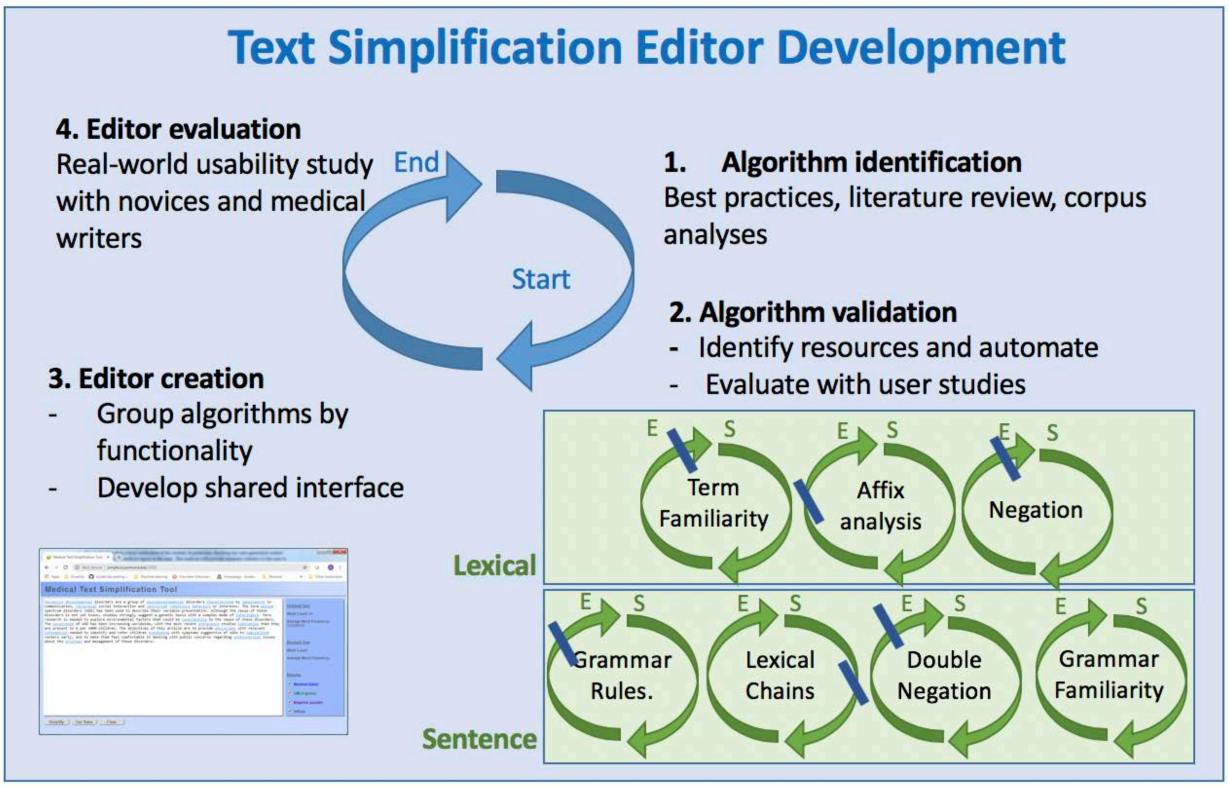

Figure 1: Overview of the Editor Creation Process.

<u>Tradeoffs and Design Choices</u>

While developing the algorithms, we kept several design choices in mind. First, we prioritized algorithms that can clearly identify difficult sections in text and provide either simpler alternatives or concrete guidance to the writer. Simply highlighting difficult sections without providing an evidence-based easier alternative is not an effective approach to improving text since our target audience for the tool is all writers of medical text, without any educational requirements, e.g. in linguistics or other related fields. For example, our grammar familiarity study has categorized thousands of different high-level grammar structures based on their difficulty [6], however, it is difficult to automatically suggest simpler grammatical structures for each possible sentence and so they have not been integrated into the tool yet. In contrast, using a parallel corpus of easy and difficult text, we have identified approximately 140 grammatical rules that both identify difficult sentences and then give concrete guidance on how to simplify that sentence using easy to understand descriptions and example sentences showing the rule application.

Second, we do not want to overwhelm the tool user with too many simplification suggestions. Therefore, we have prioritized the algorithms and suggest those with the highest impact on simplification first. For example, the tool identifies difficult terms and gives candidate suggestions. When we find a difficult term, if we can give simpler candidates in our lexical resources (i.e. WordNet[7] and UMLS[8]), then we present those candidates. Only if no candidates exist from these resources do we show the affix analysis results, which still provide guidance, but are a bit noisier. As another example, many of the algorithms have parameters that we can adjust to vary what level of difficulty warrants flagging. We have been careful to adjust these parameters to balance between being effective without overwhelming.

Tool Functionality

Figure 2 shows a first version of our prototype which integrated lexical simplification algorithms. The text is parsed and terms that are identified as difficult (based on term familiarity scores) are highlighted in the text. Initial summary statistics of the text are displayed on the right. The simpler versions of individual terms are generated by the lexical algorithms. Writers may click on a highlighted word and different color-coded options are displayed to indicate the source of the simplified term. The writer may select which resources it would like to use for these suggestions by selecting/deselecting the options in the bottom right. For example, the term ‘pervasive’ is considered difficult because its term familiarity and one easier alternative term is available, i.e., ‘widespread’. Writers only need to click on the term to have it replace the original one in the text. In addition to selecting options, the writer can edit the text freely to make appropriate adjustments as suggestions are selected. Once the writer is done editing the text, the writer can get final statistics about their document, including a comparison with the original text.

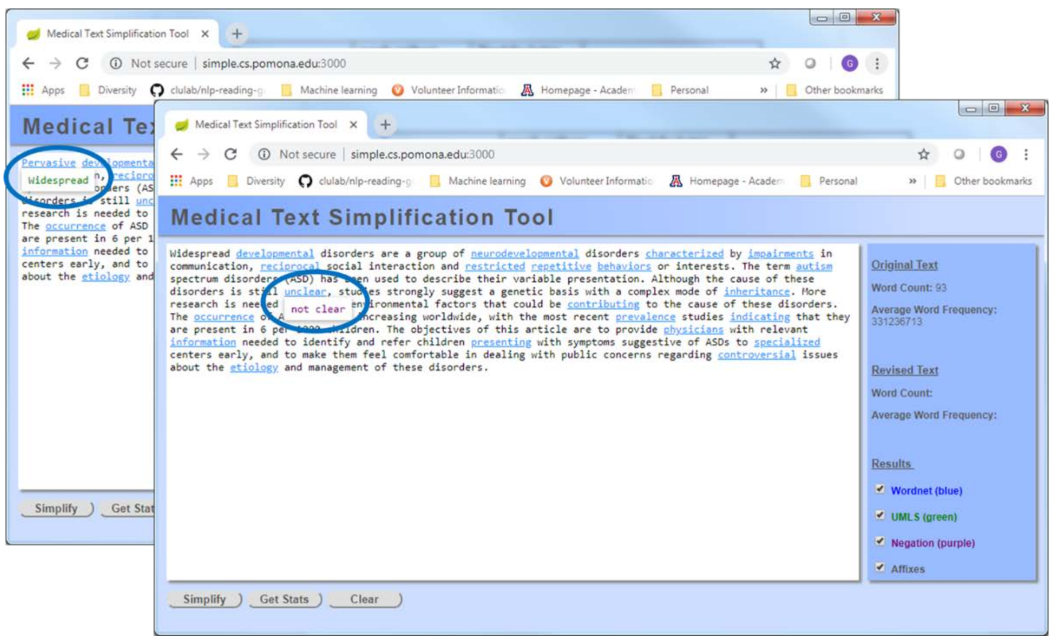

Figure 2: Screenshot Lexical Simplification

Demonstration

Table 1 gives an overview of the algorithms and their status. We propose to demonstrate our system with five algorithms: three that suggest lexical simplifications and two that focus on sentence level simplification. The demonstration will be on a live, working version of the tool that is publicly available.

Table 1: Overview of Algorithms Shown to Affect Text Difficulty

Algorithm	Purpose	Status	Language
Lexical Level			
Term Familiarity[9]	Evaluated complexity of terms, suggest easier synonyms using WordNet[7] and the UMLS[8]	Completed.	English/Spanish
Negation [10]	Identify sentential, morphological and double negations and suggest easier alternative.	Completed.	English
Affix analysis [11]	Explanation generation based on morphological analysis for words without easier synonyms	Integration Completed. Requires improvement in text presentation.	English/Spanish
Noun Phrase Splitting [12]	Identify when to split noun phrases (which is not as common as usually thought)	Algorithm ready for integration.	English
Sentence Level			
Grammar Familiarity [6]	Identification of difficulty levels using parse tree complexity	Algorithm ready for integration.	English
Grammar Simplification Rules	Generation of grammar transformation rules.	Completed. Requires improvement of rule explanation	English
Parenthesis Insertion [13]	Optimal use of parenthesis to provide different types of explanations in a sentence	Algorithm under development.	English
Document Level			
Lexical Chains	Identify dense sections in text indicative of higher difficulty	Algorithm ready for integration.	English/Spanish

Conclusion

We believe our editor has significance both from a research perspective as well as a practical. For research, we will collect usage data and use this as implicit feedback to evaluate different algorithms and resources. When the text collection becomes large enough, it will serve as a corpus for training of machine learning algorithms to learn to improve the tool as well as for general text simplification. Our editor also has practical uses and we will provide it publicly as the first text simplification tool based on extensive user testing.

Our future work will include development of additional algorithms as well as the including the Spanish version. Furthermore, to ensure usability, we will work with health educator volunteers from local community health centers to evaluate the tool. There is much evidence that five testers commonly discover 55-85% of all problems[14, 15]. Given our simple interface, we believe a similar sized group will suffice. Once we have publicized our tool, we will track use and application of algorithms to create further training data to improve the editor.

Our editor is currently available online (http://simple.cs.pomona.edu:3000/) and two individual algorithms are also available for download (NegAIT: https://github.com/kloehnen/NegAIT , Affix analysis: https://github.com/kloehnen/SubSimplifyEnglish)

Acknowledgement

Research reported in this paper was supported by the National Library of Medicine of the National Institutes of Health under Award Number R01LM011975. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] S. Fox, "Health Topics," Pew Research Center's Internet & American Life Project, Washington DC2011.

[2] Y. Wang and Z. Liu, "Automatic Detecting Indicators for Quality of Health Information on the Web," *International Journal of Medical Informatics,* vol. 76, pp. 575-582, 2007.

[3] M. A. F. Kirkpatrick and C. P. Mohler, "Using the Readability Assessment Instrument to Evaluate Patient Medication Leaflets," *Drug Information Journal,*, vol. 33, pp. 557-563, 1999.

[4] P. Norvig, "Inference in Text Understanding," presented at the National Conference on Artificial Intelligence AAAI-87, Seattle, Washington, 1987.

[5] A. C. Graesser, D. S. McNamara, and J. M. Kulikowich, "Coh-Metrix : Providing Multilevel Analyses of Text Characteristics," *EDUCATIONAL RESEARCHER*, vol. 40, pp. 223-224, 2011.

[6] D. Kauchak, G. Leroy, and A. Hogue, "Measuring Text Difficulty Using Parse-Tree Frequency," *Journal of American Society of Information Science and Technology (JASIST),* vol. 68, 2017.

[7] G. A. Miller, R. Beckwidth, C. Fellbaum, D. Gross, and K. Miller. (1998). *Introduction to WordNet: An On-line Lexical Database*. Available: http://www.cogsci.princeton.edu/~wn

[8] A. McCray, "Representing biomedical knowledge in the UMLS Semantic Network," in *High-Performance Medical Libraries: Advances in Information Management for the Virtual Era*, B. NC, Ed., ed Westport, CT: Meckler Publishing, 1993, pp. 45-55.

[9] 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),* vol. 15, p. e144 (doi:10.2196/jmir.2569), 2013.

[10] P. Mukherjee, G. Leroy, D. Kauchak, S. Rajnarayanan, D. Diaz, N. Yuan, T. Pritchard, and S. Colina, "NegAIT: A New Parser for Medical Text Simplification Using Morphological, Sentential and Double Negation," *ournal of Biomedical Informatics,* vol. 69, pp. 55–62, 2017.

[11] 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 (JMIR)*, vol. 20, 2018.

[12] G. Leroy, D. Kauchak, and A. Hogue, "Effects of Text Simplification: Evalution of Splitting up Noun Phrases," *Journal of Health Communication: International Perspectives,* vol. 21, pp. 18-26, 2016.

[13] Y. Gu, G. Leroy, and D. Kauchak, "When synonyms are not enough: Optimal parenthetical insertion for text simplification," presented at the AMIA Fall Symposium, Washington DC, 2017.

[14] L. S. Zun, T. Sadoun, and L. Downey, "English-language competency of self-declared English-speaking Hispanic patients using written tests of health literacy," *Journal of the National Medical Association,* vol. 98, pp. 912-917, June 2006.

[15] J. Nielsen, *Usability Engineering*: Morgan Kaufmann, 1993.