SimplyPut: Leveraging a Mixed-Expertise Crowd to Improve Health Literacy

Edith Law[1], Barbara Grosz[1], Lee M. Sanders, M.D.[2], and Shira H. Fischer[3]

[1] Center for Research on Computation and Society
SEAS, Harvard University
elaw@seas.harvard.edu,grosz@eecs.harvard.edu
[2] Division of General Pediatrics
Stanford University
leesanders@stanford.edu
[3] Division of Clinical Informatics
Harvard Medical School
shira@post.harvard.edu

Abstract. A substantial amount of health information, online or in print, is much too technical for the general patient population to understand. Our project aims to create a system that combines intelligent agents with a mixed-expertise crowd to translate medical documents into text that can be digested and interpreted correctly by parents of chronically ill children.

Keywords: Interactive Human Computation, Health Literacy, Health Communication, Mixed-Expertise Crowd, Crowdware, Text Simplification and Summarization

1 Introduction

Parents are now able to draw on a wealth of information on the Web to help them manage the health of their children. For almost any disease, a variety of websites, wikis, and forums describe such key disease factors as causes, symptoms, prognosis, methods of treatment and prevention. This information is, however, not accessible to a large segment of the population: some have low literacy, others are unable to interpret correctly the results of studies, and yet others cannot determine the meaning of a general study for their individual child's care. A recent study [14] evaluating the health literacy of parents reveals that almost 30% of parents have below basic or basic level of health literacy. In general, about 20 percent of adult Americans have 5th grade or lower reading level [3], which is far below the level required to understand most print or online medical and healthcare literature. Low parental health-literacy levels are correlated with an inability to adhere to instructions and take appropriate actions, which ultimately negatively affects the health outcomes of their children. Similar challenges are faced by adult patients with lower reading level managing their own health.

To be useful, health information should be both *readable* and *comprehensible.* The readability of a piece of text refers to its difficulty, as measured by properties of the text itself, such as vocabulary level, sentence structure, use of technical jargon, and density of information [7]. In contrast, comprehension measures the extent to which the reader can accurately remember and interpret the information contained in the text. Research in natural-language processing has yielded a variety of methods for text summarization [10], simplification, sentence compression [11], and the prediction of reading levels, some of which have been focused on medical documents [4]. None of these methods have yet been used to transform medical documents at scale, and to-date the techniques have not been able to provide the accuracy and completeness needed for use by patients.

This paper describes research aimed at developing a collaborative multi-agent "human computation" system, called **SimplyPut**, which coordinates people with different levels of medical expertise, along with text-processing algorithms, to address the problem of producing healthcare information appropriate for readers at lower levels of health literacy. This effort builds on our prior work on "crowdware" [15]—applications designed to coordinate a crowd involved in a common task to achieve a goal—and on collaborative multi-agents systems. This system will coordinate a crowd of workers to perform the task of translating paragraphs from a medical journal article into simpler, more compact text that can be more easily understood by those with lower literacy levels. We treat this transformation as a kind of machine translation problem, in that it takes information written in technical medical language and translates it into simple, lay language that conveys the important points for patients (or their parents). This approach is similar to Soylent [2], which deploys crowdsourcing for editing and shortening text.

Ideally, one would be able to find workers who are *bilingual* in the sense that they know both the language of medicine and how to "speak plainly". There are insufficient people with such skills to address the problem at scale. Instead, we proposal a solution involving a *mixed-expertise crowd* comprising non-expert workers (e.g., Mechanical Turk workers, health forum members), semi-experts (e.g., medical students), and experts (e.g., doctors). This team works together, iteratively refining drafts to produce a final translation. Our goal is to develop a system that coordinates human agents from the mixed-expertise crowd and computer agents (e.g., NLP algorithms), intelligently dividing the labor between the two based on their estimated competence at various text transformation tasks. Different from Soylent, our system encourages the active involvement of the requester in guiding the workers and in shaping the final solution. Involving the requester (i.e., physician) is particularly important in the medical domain, where the translated material must not only be accurate and complete, but also serve the particular needs of the patients and their caretakers, making them aware not only of what the material says, but also of what it *means* for them.

This work is part of a broader project that aims to develop intelligent, autonomous multi-agent systems that work as a team supporting a diverse, evolving team of providers caring for children with complex conditions.[4] In this paper,

[4] This project accords with the vision described in the paper Collaborative Health Care Plan Support in the challenges and visions track of the technical conference.

we describe our vision for this research; specifically, we characterize the health literacy problem in greater detail, present observations from a pilot experiment, describe the initial system design, and discuss the anticipated challenges.

2 Characterizing the Problem

To understand the nature of the problem and the desired solution, we conducted a pilot experiment in which we examined differences between experts and non-experts in simplifying medical documents. The study used the following paragraph from a medical journal about the effects of oral corticosteroids on growth for kids with asthma.

> There is little doubt that oral corticosteroids such as prednisolone can have a detrimental effect on growth. Martin et al, in a prospective survey over 14 years, showed that children who had received oral steroids were significantly shorter than either asthmatic children who had not received steroids or non-asthmatic controls. However, this difference in height was only seen at age 14 years, and no difference was apparent by 21 years, indicating that the main effect of oral corticosteroids was to cause growth delay and affect the timing of puberty. The degree of growth retardation has been clearly linked to the frequency of oral corticosteroid use. However, there is also evidence that adult height can be permanently reduced in some children who have received long term oral corticosteroids for asthma.

Four medical doctors, with different types of expertise (neurology, pediatrics, general medicine) were asked to generate a 2-3 sentence summary that was appropriate for conveying to patients the information contained in the paragraph. They were asked also to articulate the process they used to generate the summary. The results are presented in Table 2.

Even in this small study a number of interesting factors emerged: different doctors assumed different specific goals. Some doctors focused on simple reporting of the facts: their process of translation involves simplification of the language, such as explicit steps to remove medical jargon, remove words with too many syllables, reduce the total number of words in the sentence and the density of the information in the text. Others added extra information, such as the definition of oral corticosteroids, to clarify the statements. One doctor went as far as constructing a "take home" message based on the paragraph, by developing two broad categories – "what does this say" and "what does this mean for my child" – and sorting the information in the paragraph into those categories.

Despite their differences, the doctors all clearly adopted a patient-centric approach to designing their summaries. They considered "what might confuse the patient," "[what] words have different meanings in medicine than in normal parlance," "what the goal was for the parents to know," "translating subjects / controls into what this means for their kids," or substituting in terms that are more medically relevant, such as "side effect" and "dosage."

Translation	**Process**
A kind of medicine called "oral corticosteroids" that we use for asthma can negatively affect a child's growth. One study watching children's development over 14 years found that children on this medicine were shorter at age 14 than children who did not receive it. However, this was only at age 14: by the time these children were 21 years old, their height was normal, so the medicine may have just slowed their growth when they were younger and made puberty later. The more medicine the took, the stronger this effect was. However, there are other studies showing that perhaps the medicine may have permanent effects on height if they've taken the medicine for a long time.	I think there's some gestalt, like what about this might confuse the patient, versus the specifics. I am sensitive to words that have different meanings in medicine than in normal parlance, so however well you explain the study, I think it'll still be confusing if the patient doesn't know what steroids are, for example. I read the whole thing, thought about it, and then did the summary. I also thought about what the goal was for the parents to know - that a) there are side effects and b) the research is unclear about whether these are short-term or long-term. That's not very reassuring, but that seemed to be what we had, so that was all we could tell them, but I wanted at least that much to be clear.
Steroid pills, like the kind that children with bad asthma sometimes need to take, carry some bad side effects. One side effect is that kids are shorter, or may go through their growth spurt or puberty later. If only on steroids occasionally, the kids usually catch up, but if the steroid course is prolonged they may be shorter as adults.	Simplify language. Set the discussion up as a 'side effect' of treatment. Switch from population level discussion to one aimed more at a parent with a child, translating subjects/controls into what this means for their kid.
There is some evidence that children who take oral steroids (a class of anti-inflammatory medications) for asthma may experience some slowing of their growth. However, studies disagree on whether this effect is permanent, or whether these children eventually catch up. The frequency of use may be important.	I first read the entire paragraph, which contained sentences presenting several at times conflicting facts about steroids and growth, and identified the "lowest common denominator" with which all the sentences were in agreement, which is that steroids seem to have some effect on growth. I made this the topic of the first sentence in my summary and rewrote this to eliminate jargon. Next, I identified where the various sentences were in disagreement (which is regarding whether effects on growth were temporary or permanent) and made this the second sentence in my summary. Finally, I focused on the one sentence that may account for the discrepancy, which is the one concerning dose effects.
What does this say? If a child takes steroid pills or liquid by mouth, it may cause her to be shorter when she is an adult. This is mainly true for a child who takes steroid pills every day for one month or more. It is also more of a problem for children who take steroids before they are 14 years old. **What does this mean for my child?** If your child needs to take steroid pills for more than one month in a single year, ask her doctor if it may affect her growth. Your child may also need to use a steroid inhaler (puffer). A steroid inhaler is safe. It will not cause her to be shorter when she is an adult. If she does not take the steroid inhaler, she may get very sick.	I followed clear-health communication guidelines (Doak and Doak): - Answer the simplest and most actionable question(s). No more than 3 questions. - Choose no more than 3 messages per question. - Use the fewest possible sentences and the fewest possible words. - Use words with $<$ 3 syllables. Avoid any medical jargon. - Use sentences with $<$ 12 words. - Avoid phrases and complex sentences. - Avoid commas and parentheses. - Use large font size and bullets. - Maximize white space. I added my clinical experience.

Table 1. Doctors' Translations and Process

Translation	**Process**
A new survey showed that children who received oral steroids were shorter than other children. However, they caught up by age 21, showing that there was a delay in growth and not a true stunting. Long-term use may still stunt growth.	I tried to read a few sentences, then summarize as plainly as possible what the article was trying to say.
There is no proof that steroids administered orally has any detrimental effect on growth. It may stall growth at certain stages of life, thereby delaying puberty, but once the administration of steroid is stopped, growth would progress normally as before.	I had to avoid all possible medical terminology because the patient will not be understand it and so there is no point in elaborating it. Since there is no higher education, the explanation mustn't frighten the patient. It must instill faith in him. The patient must feel confident after the explanation.
Studies show that the use of oral steroid medicine, for treating asthma, can delay puberty in children. These children are shorter at or around 14 years old but eventually catch up by adult age. Although in some cases of prolonged use could cause permanent effects in height.	I just tried to make the paragraph be very simple and to the point. Avoiding difficult, medical jargon. I feel in such a situation the studies do not need to be intro as thoroughly. Just state the main problem, and then eliminate any worry by saying what is normal and can be done.
Studies show that children who use oral steroids were shorter than children who didn't at age 14, but were all the same by age 21. However, some children that used a lot of the steroids were permanently shorter than children who didn't use them as much.	I moved sentence by sentence and simplified terminology and combined sentences where possible.
Certain surveys do show a significant decrease in height in children who take oral steroids for asthma,but such difference in height was only seen at age 14 years, and no difference was apparent by 21 years. But there is no evidence yet to proves that adult height permanently reduces in children who take oral corticosteroids for asthma for a long time.	Went through the whole para first,understood in a normal manner and then typed the jist

Table 2. Turkers' Translations and Process

We ran an equivalent task on Mechanical Turk, in which we asked five workers with no medical expertise to summarize the paragraph in 2-3 sentences as if they were talking to a patient (who has low reading level) about it, and describe the process or steps they took to generate the summary. In general, Turkers seemed to have produced more straight-forward summaries that are sentence-by-sentence translations, involving less re-organizing of content and having fewer elaborations; very few translations address what the information *means* for the patient. Some translations are missing important facts; for example, the second translation failed to report the fact that prolonged usage of corticosteroids can have a permanent effect on height. None of the Turkers substituted in terms that are more familiar to patients, such as "dosage" and "side effect." Our vision is to create a system that enables the requester, e.g., a doctor, to iteratively point out such errors and suggest ways to rephrase the text to be more patient friendly.

3 SimplyPut: A Crowdware System for Translating Technical Documents

We present here our vision of a system called *SimplyPut*, a web application where experts and non-experts (e.g., Mechanical Turk workers) collaborate using a shared interface to translate technical documents for a target audience. In the medical domain, experts are medical professionals, the documents are medical information from journals or websites, and the audience are patients or parents of patients. The system will use a novel approach that leverages a mixed-expertise crowd and will support human-computer collaborations in this context.

To date, most human computation systems have been developed to address problems that allow for a task to be decomposed into operations that can be performed by individual workers, each working independently and having only a local view of the solution. A different approach is required if a task cannot be fully decomposed into individual tasks, i.e., if there are dependencies among the tasks undertaken by different workers. We have developed such an approach in prior work on a crowd-based collaborative interface called Mobi [15], in which workers asynchronously put together a travel itinerary. In Mobi, workers choose the type and amount of contributions they want to make, while the system exerts some control over the computational process by automatically tracking the violated constraints and alerting workers about things that need to be adjusted to meet these constraints. Mobi is essentially a groupware system for crowds – referred to as crowdware. It leverages communication, a shared interface, and mechanisms for keeping track of the goal and progression towards it, to help a group to accomplish a joint task. The underlying idea behind crowdware is an **interactive approach to human computation**: requesters and workers are given the ability to monitor the solution as it is being assembled and influence the computational outcome through intermittent communication with one another. The goal of *SimplyPut* is to allow the requester (e.g., a doctor) to interact with and guide workers with little or no medical expertise to produce a final document that satisfies his or her requirements.

Like Mobi, *SimplyPut* will incorporate a shared interface that different agents, whether human agents or computer agents, will use to collaborate and generate the solution together. By allowing human agents to critique other human agents and computer agents, the system can then adaptively decide on the division of labor [12, 1, 6] between human and computer, assigning to each party tasks at which it is most competent. Getting the division of labor right is a key requirement for collaborative interfaces [1]. Having expert and semi-expert workers amongst the crowd, the system can guarantee that the final document serves its true purpose, which in this case is to communicate to patients and their caretakers what they need to know, no more and no less. Finally, by keeping the requester in the loop, allowing them to monitor the state of the solution and influence the final outcome, we can ensure, to some extent, that the final translation is accurate and complete. We now describe the different aspects of the system and some key challenges.

3.1 Agents

In our system, there will be three different kinds of agents:

Requesters are human agents who are the *users* of the system needing a technical document to be translated.

Workers are human agents responsible for carrying out various text processing and verification tasks, including non-experts with no formal medical training, or semi-experts (e.g., medical students) and experts (e.g., doctors).

NLP Algorithms are computer agents responsible for carrying out various text processing and verification tasks.

We envision that these agents will interact in a variety of ways. The role of the requester is mainly to provide requirements and constraints to the worker agents, and to critique and provide feedback on their work. Human workers transform the text, by adding, removing, re-ordering content, and also providing feedback for other human or automated agents. Automated algorithms can perform text modification operations. Given evaluative feedback from the requester and other human worker agents, these algorithms can also update their behavior by learning from past mistakes or additional training examples. For example, the system can send the output of a sentence compression algorithm to human workers, asking them to identify and correct mistakes, e.g., if the algorithm has mistakenly deleted parts of the sentence that are important, or if it has failed to identify superfluous parts of the sentence. This evaluative feedback from the human worker agents can then serve as new training example for the sentence compression algorithm.

A key challenge in designing this system will be to get the division of labor right, assigning to the computer agents only tasks they can do well, and likewise for workers at lower levels of the expertise hierarchy. We will follow an iterative, learning approach to this problem. Initially we will assign to both computer agents and people the same tasks—e.g., take a sentence and remove unnecessary elements from it—and then have human agents vote on the best results. If a sentence compression algorithm is comparable to human performance, then the system should leverage it over human workers; otherwise, the system will delegate the tasks to human workers.

3.2 Tasks

There are a variety of text transformation tasks that human workers and automated algorithms can perform. Some of these tasks are on the sentence level, while others may be on the paragraph level. They will include

Highlight Identify words or phrases that can be simplified (e.g., jargons).
Substitute Replace a word or phrase with easier to understand terms.
Chop Compress text to remove unimportant words or phrases.
Re-Order Reorder parts of the text (e.g., changing passive to active voice).
Elaborate Add definitions or explanations to the text.
Condense Combine multiple paragraphs by retaining only the important facts.
Verify Judge the work of other human or computer worker agents.

These tasks can be combined into a workflow, and assigned to workers by the system or by the requester. In particular, the requester can monitor the current solution, identify problematic areas, and assign new text transformation tasks to fix the problems. For example, the requestor may realize that he wants the sentences to be shorter, in which case, he may assign a *chop* task with the requirement that each sentence be limited to at most n words.

3.3 Interface

Figure 1 illustrates our initial design of SimplyPut. In the middle panel, the system presents to the worker a paragraph and a specific text transformation task. Users can choose to work on the current paragraph, or skip to some other paragraphs (and tasks) by scrolling up or down.

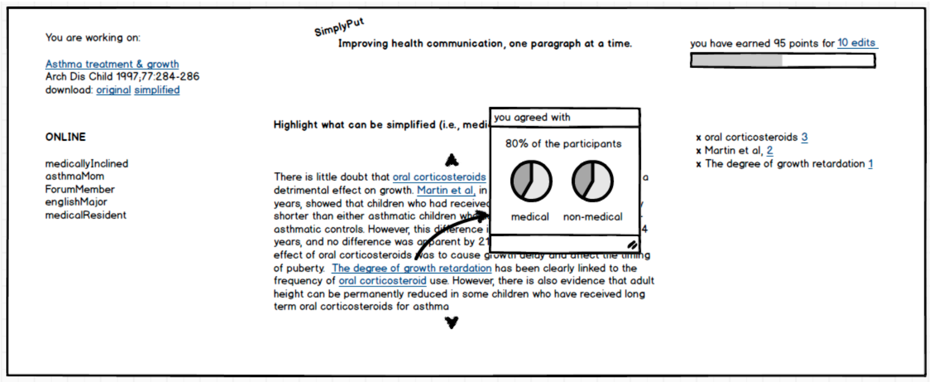

Fig. 1. SimplyPut: An Envisioned Design

The example shown here is a *highlighting* task, which asks the worker to select words or phrases in the text that can be simplified. With each selection, the system provides immediate feedback to the worker, indicating whether his or her selection agrees with what other participants have previously chosen. The worker can download the original article as well as the (partially or completely) simplified version. Finally, the interface displays information about the progress of the translation and the amount of contribution the worker has made.

3.4 Challenges

Evaluation Translations produced by our system should be correct, readable and comprehensible to the intended audience. We can evaluate each translation on its perceived and actual readability [8] using both automated metrics (e.g., Flesch [5] and SMOG [9]) and human judgments. To ensure that our tool is useful in practice, we will also evaluate the comprehensibility of the translations in a variety of ways. A common method for evaluating comprehension is the Cloze procedure [13], which asks evaluators to fill in the blanks of a written passage which has words deleted at a randomly chosen set interval. Alternatively, one can evaluate comprehension using explicit questions. We can, for example, ask the requesting doctor to generate a list of take-home facts that a patient should remember about the text, as well as a list of questions that will test the comprehension of these take-home facts. Each translation will be first reviewed to see how many and which of the take-home facts they retained (which measures its precision and recall), then evaluated on its comprehensibility by asking human evaluators to answer questions (generated by the requesting doctor) about those take-home facts.

Incentives Our system is aimed at bringing together a variety of people—doctors, medical and health communication students, as well as interested citizens (e.g., members of various health forums)—to collaboratively simplify medical documents. While doctors may be interested in disseminating scientific research findings to their patients in an understandable way, other participants may be interested in reading about certain health conditions and diseases, meeting other people with similar interests, educating themselves on how to write in a patient-friendly language, or contributing to the greater good of improving health literacy and communication. A major challenge, therefore, is in designing our system to simultaneously address a variety of intrinsic motivations.

Division of Labor Non-experts may produce lower quality work, but have substantially more time to devote to the endeavour than semi-experts or experts. Natural language processing algorithms may be less competent, but significantly cheaper to deploy than paying humans to perform the same tasks. To be cost-effective, our system needs to optimally divide the work amongst non-experts, experts and automated algorithms to produce translations, while taking into account monetary and time constraints.

4 Conclusion

In this paper, we present the health literacy problem and propose to address it though development of a collaborative crowdware system called SimplyPut. The human and computer agents in this system will collaborate to translate medical documents, making them more accessible and easy to understand for the general patient population. Our system is intended to be as much of a communication vehicle as it is a translation tool. Through this large-scale, collaborative effort to simplify medical information, we hope that people from otherwise disjoint communities—i.e., doctors, students, patients—will begin to understand how to better communicate with one another.

Acknowledgments. We thank Stuart Shieber for numerous discussions about the direction of this project, and for his valuable suggestions and feedback. The research reported in this paper was supported in part by grant IIS-0705406 from the U.S. National Science Foundation.

References

1. T. Babaian, B. Grosz, and S. Shieber. A writer's collaborative aid. In *UIST*, pages 7–14, 2002.
2. M. S. Bernstein, G. Little, R. Miller, B. Hartmann, M. Ackerman, D. Karger, D. Crowell, and K. Panovich. Soylent: a word processor with a crowd inside. In *UIST*, pages 313–322, 2010.
3. C. Doak, L. Doak, and J. Root. *Teaching Patients with Low Literacy Skills*. J.B. Lippincott Company, 1996.
4. N. Elhadad. *User-Sensitive Text Summarization: Application to the Medical Domain*. PhD thesis, Columbia University, January 2006.
5. R. Flesch. A new readability yardstick. *Journal of Applied Psychology*, 32:221–233, 1948.
6. B. Grosz and L. Hunsberger. The dynamics of intentions in collaborative intentionality. *Cognitive Systems Research: Special Issue on Cognition, Joint Action and Collective Intentionality*, 7:2–3, 2005.
7. G. Leroy, S. Helmreich, and J. Cowie. The influence of text characteristics on perceived and actual difficulty of health information. *International Journal of Medical Informatics*, 79:438–449, 2010.
8. P. Ley and T. Florio. The use of readability formulas in health care. *Psychology, Health and Medicine*, 1(1):7–28, 1996.
9. G. McLaughlin. Smog grading – a new readability formula. *Journal of Reading*, 12(8):639–646, 1969.
10. A. Nenkova and K. McKeown. Automatic summarization. *Foundations and Trends in Information Retrieval*, 5(2-3):103–233, 2011.
11. E. Pitler. Methods for sentence compression. Technical report, University of Pennsylvania, 2010.
12. S. M. Shieber. A call for collaborative interfaces. *ACM Computing Surveys*, 28(4es), 1996.
13. W.L. Taylor. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30:415–433, 1953.
14. H.S. Yin, M. Johnson, A.L. Mendelsohn, M.A. Abraham, L.M. Sanders, and B. P. Dreyer. The health literacy of parents in the united states: a nationally representative study. *Pediatrics*, 124(3):289–298, 2009.
15. H. Zhang, E. Law, K. Gajos, E. Horvitz, R. C. Miller, and D. Parkes. Human computation tasks with global constraints: A case study. In *CHI*, 2012.