Optimizing the *txt2MEDLINE* Search Portal for Low-Resource Clinical Decision Support

Lincoln Sheets MD MS-CIS¹,², Fang Liu MS², Raymond Francis Sarmiento MD², Alex Gavino MD², Paul Fontelo MD MPH²

¹University of Missouri Informatics Institute, Columbia MO
²National Library of Medicine, Bethesda MD

Abstract

txt2MEDLINE brings high-quality medical evidence via text-messaging to settings with inadequate Internet access. We optimized the txt2MEDLINE search technique by parsing queries for MeSH (Medical Subject Heading) terms and searching MEDLINE for articles containing these terms in their titles or abstracts. We compared our results to the existing txt2MEDLINE tool by compiling benchmark queries from low-income and low-middle-income countries, and asking doctors and nurses with practice experience in low-resource areas to evaluate them. The median scores on a 5-point Likert scale were 2.9 for the existing txt2MEDLINE vs. 3.8 for the modified version (p=0.015). This reached our predefined criterion for clinical significance, a difference of 0.5 standard deviations. Improving this technology could improve clinical information resources in the world’s most medically underserved communities.

Introduction

The Internet and smartphones are increasingly important for clinical decision support, but both are limited by cost or coverage in many medically underserved settings. txt2MEDLINE surmounts this problem by providing access to high-quality medical evidence via text-messaging. However, these search results are in reverse chronological order rather than order by best match, which may make it less useful for point-of-care decision support. We developed an optimized txt2MEDLINE search technique to produce query results more useful for answering clinical queries by SMS in low-resource settings.

Background

In the United States, a looming expansion in Medicaid coverage and an expected shortage of primary-care physicians is expected to exacerbate existing problems of healthcare access¹. The problem is even greater in the world’s most medically underserved regions, where the density of physicians per capita is often less than one-tenth that of most industrialized countries². US agencies and legislatures will consider expanding the scope of care for advanced-practice nurses and other healthcare providers, a strategy already used by necessity in the rural regions of many developing countries³; yet a knowledge gap will remain between these dedicated non-physician providers and a standard-of-care Western medical education.

For trained healthcare workers with experience in taking patient histories and conducting physical exams, the Internet could bridge that gap⁴. However, an economic disparity in access to the World Wide Web parallels the disparities in healthcare access and healthcare workforce development. In 2010, 95% of Americans making over $75,000 per year used the Internet, compared to 57% of those making less than $30,000 per year⁵. This “digital divide” disproportionately affects minorities, with only 51% of African-American adults owning a desktop computer compared with 65% of whites⁶. While 87% of African-Americans and English-speaking Hispanic Americans owned cell phones, even more than the 80% of white Americans who did⁷, those phones or carrier contracts limited to voice and short messaging service (SMS) texting cannot access the resources of the World Wide Web. Around the world, and increasingly in the United States, the healthcare of billions is in the hands of non-physician providers who are dedicated and experienced but not educated as physicians; who are literate and intelligent but can’t afford the most modern electronic tools; and who talk and text on cell phones every day but cannot use those phones to access the medical resources of the World Wide Web.

The US National Library of Medicine developed txt2MEDLINE⁸ precisely to bridge this gap by providing access to high-quality medical evidence via text-messaging. While past research into automated SMS interfaces for healthcare has focused on monitoring and modifying patient knowledge and behavior⁹, txt2MEDLINE appears to be the only
tool using SMS interfaces for point-of-care information retrieval by healthcare providers. Previous studies have shown strengths and limitations of this approach. Since txt2MEDLINE operates within the restricted screen space of text-messaging, its users may depend much more on the relevance of the first one or two results, rather than browsing through many results as on a computer screen. Because results are in reverse chronological order, however, the chances are relatively high for any given query that the first result will be a narrowly focused laboratory finding that is more research-oriented and less pertinent to clinical practice. The first published example query for txt2MEDLINE, “Carpal tunnel syndrome, Surgery versus steroid injection, Randomized controlled trial,” returned and still returns only one result. For queries with many results, however, optimizing the usability of the first result for each search has the potential to greatly increase the tool’s effectiveness for clinical decision support.

Our objective was to compare an optimized txt2MEDLINE search technique to the existing txt2MEDLINE tool. Our hypothesis was that optimizing search strategies for txt2MEDLINE would produce query results significantly more useful for answering clinical queries by SMS in low-resource settings. We defined a clinically significant improvement as an increase of at least 0.5 standard deviations over the existing result.

Methods

Modification of txt2MEDLINE
We added processing to txt2MEDLINE to parse queries for MeSH (Medical Subject Heading) terms and search MEDLINE for articles containing these terms in their titles or abstracts. Articles without abstracts in MEDLINE were excluded. Abstracts with some or all of the queried MeSH terms in their titles were ranked higher than those with the terms in their abstracts only, and reviews were ranked higher than other publication types. The first 40 results were scored based on the number of unique and repeated uses of the queried MeSH terms in the titles and computer-generated summaries, with penalties for other MeSH terms, the word “withdrawn” or the words “new,” “novel,” “emerging,” or “recent,” in order to avoid research findings with untested clinical applications. The result with the most points was selected (Figure 1).

![Figure 1: Algorithm for the modified txt2MEDLINE.](image-url)
Development of Benchmark Queries

A total of 12,666 queries from 123 countries were made through txt2MEDLINE, askMEDLINE and PubMed PICO in 2011. Selecting only queries from low-income to low-middle-income countries (as defined by the World Bank) yielded 7,759 queries from 36 countries. We used World Bank data\textsuperscript{13} to determine the ratio of cell phones to broadband connections by country and included queries from the 12 countries where this ratio was highest, in order to target the clinical information needs of the regions least likely to use a Web browser at the point of care but likely to be able to use a text-messaging tool. These countries were Iraq, Kenya, Cote D'Ivoire, Sudan, Zambia, Bangladesh, Mali, Mozambique, Nicaragua, Honduras, Senegal, and Indonesia. A total of 625 queries from these countries were made in 2011. In order to reduce the number of queries to a manageable size, we sampled 3\% of the 625 queries (every 33\textsuperscript{rd} query in chronological order, beginning with the first query from each country in order to represent all the selected countries with at least one query). In the few cases where this method produced sample queries that were not clinically oriented, the next query from that country was used. Queries in languages other than English were translated using BabelMeSH\textsuperscript{14} (Figure 2).

Figure 2: Benchmark queries used to compare the two versions of the search tool. A total of 20 benchmark queries were extracted from the total number of queries submitted during the year 2011 to txt2MEDLINE or its sister search portals, askMEDLINE and PubMed PICO.

Comparison of Existing txt2MEDLINE and the Optimized txt2MEDLINE

We ran both versions of txt2MEDLINE against these benchmark queries and recorded the first result for each, masking the sources and alternating the order to prevent bias. Nine doctors and nurses with practice experience in low-resource areas were recruited to score the results from 1 (not useful) to 5 (very useful). All of these volunteer reviewers were practicing or former clinicians from low-middle-income countries (mostly the Philippines), although four of the nine had emigrated to high-income countries before taking the survey. We made the survey and a
paragraph of instructions available anonymously on a secure Web page, each volunteer was asked to rate all results, and we excluded any incomplete surveys.

Since the Likert-scale responses could not be assumed to be normal, we considered the median for each search portal rather than its mean, and looked for statistical significance with a non-parametric Wilcoxon T-test$^{15,16}$. With no consensus in the literature on what constitutes clinical significance for a Likert scale in this area, we followed the precedent of one-half standard deviation above the control value$^{17}$.

**Results**

The average scores for the modified txr2MEDLINE results to all twenty queries ranged from 2.6 by the lowest-scoring of the nine reviewers to 4.5 by the highest-scoring (with 1 = not useful to 5 = very useful). The range of average scores was 2.3 to 3.3 for the results to the same queries in the existing version.

The median score with standard deviation was 3.8 ± 0.66 for the modified txr2MEDLINE vs. 2.9 ± 0.34 for the existing version. The difference was statistically significant with a p-value of 0.015. The modified algorithm’s median score was more than 0.5 standard deviations greater than the median score for the existing algorithm (i.e., 3.8 – (0.5 * 0.66) = 3.47 > 2.9), exceeding our minimum target effect size and our pre-defined threshold for clinical significance.

**Discussion**

Our goal in conducting this study was to demonstrably improve the clinical usefulness of the first-returned txr2MEDLINE result for real-world queries. Our modifications to txr2MEDLINE produced results that were statistically and clinically improved, using the most objective measure of clinical significance that we could document.

In order to measure the usefulness of txr2MEDLINE before and after our changes, in the environments where we believe it is most critically needed, we sampled queries from only the 12 low-income and low-middle-income countries with the highest ratio of cell phones to broadband connections. The size of the sample was designed to create a manageably small survey for our volunteer evaluators. Limiting the benchmark queries to these countries may limit the applicability of the results to other environments. In addition, our exclusion of some queries subjectively evaluated as not clinically oriented may affect the reproducibility of these results. And while the addition of MeSH terms to the search algorithm may have improved results for medical topics, it could adversely limit the results in other health-related field (e.g., nursing, nutrition, physical therapy, etc.) that we did not explicitly test.

We distributed our survey only to physicians and nurses with low-resource clinical experience in order to accurately judge usefulness in the context of interest, with a resulting relatively small survey population. Inter-observer variation in scores was generally narrow, with standard deviations of less than 1.0 on a scale of 1 to 5; however, many examples of outlying scores of 1 and 5 from different evaluators of the same query response show that subjectivity in scoring the query results was a limitation in our study design.

We did not investigate the ethics of basing clinical decisions on the titles and abstracts of papers for which the full article is referenced but not directly provided. However, it is the consensus of the authors (of whom four have practiced medicine in low-income or middle-income countries) that txr2MEDLINE can be a valuable addition to the limited resources available in the practice settings for which it is designed. The txr2MEDLINE interface is intended for medical professionals and has been made available to clinicians in middle-income countries for previous studies$^{11,12}$. Review of this study’s design by an internal review board was waived because it meets the human subjects research exemption #2 of the National Institutes of Health’s human-subjects protection guidelines$^{18}$.

To the extent that these results can be generalized, they support the feasibility of improving txr2MEDLINE searches for clinical problems, by returning more relevant results before more recent results. Our final algorithm was the result of several iterative improvements and cost-benefit decisions; for example, searching the articles’ MeSH metadata labels instead of their titles and abstracts would have had the advantage of leveraging the MEDLINE
indexers' expertise at summarizing the relevant concepts, but we found that it produced similar results with higher processing costs and wait times. Further investigation may refine the search algorithm proposed here.

These results support the conclusion that a modified search strategy would significantly increase the clinical usefulness of txt2MEDLINE results. This could improve clinical information resources in the world’s most medically underserved communities.

References