

# Development and Evaluation of a Prototype Search Engine to Meet Public Health Information Needs

Jonathan W. Keeling, MS<sup>1</sup>, Anne M. Turner, MD, MPH<sup>2</sup>, Eileen E. Allen, MLS<sup>3</sup>,  
Steven A. Rowe<sup>3</sup>, Jacqueline A. Merrill, RN, MPH, DNSc<sup>1</sup>,  
Elizabeth D. Liddy, PhD<sup>3</sup>, Howard R. Turtle, PhD<sup>3</sup>

<sup>1</sup>Columbia University, New York, NY; <sup>2</sup>University of Washington, Seattle, WA; <sup>3</sup>Syracuse University, Syracuse, NY

## Abstract

*Grey literature is information not available through commercial publishers. It is a sizable and valuable information source for public health (PH) practice but because documents are not formally indexed the information is difficult to locate. Public Health Information Search (PHIS) was developed to address this problem. NLP techniques were used to create informative document summaries for an extensive collection of grey literature on PH topics. The system was evaluated with PH workers using the critical incident technique in a two stage field evaluation to assess effectiveness in comparison with Google. Document summaries were found to be both helpful and accurate. Increased document collection size and enhanced result rankings improved search effectiveness from 28% to 55%. PHIS would work best in conjunction with Google or another broad coverage Web search engine when searching for documents and reports as opposed to local health data and primary disease information. PHIS could enhance both the quality and quantity of PH search results.*

## Introduction

Public health (PH) is a broad domain that provides population health services of wide scope. To ensure effective delivery of these services, the PH workforce is composed of a variety of professional positions, training and skills. The diversity of the PH workforce challenges efforts to meet information needs that are far less specific than in the clinical setting. Despite these challenges, a review published in 2007 found 3 broad areas of information need: synthesized and summarized information from databases and research reports; new content from grey literature such as policy documents and best practices; and current national, state, and local health data such as birth rates and ER visits<sup>1</sup>.

Information of this type is generally not published or indexed in the scientific literature<sup>2, 3</sup>. This information is described as “fugitive” or grey literature, i.e. literature that is difficult to locate because it is not available through traditional commercial channels<sup>4</sup>. These documents, which include best practice reports, policy documents, community surveys and standard forms, are produced by PH experts and practitioners but not published along formal publishing pathways. While much of this information is valuable, traditionally such documents remained isolated within organizations. These documents are increasingly put on the Web, yet the lack of uniformity in formats and indexing schemes continues to make access difficult. Even trained information specialists find searching for grey literature inefficient<sup>5</sup>. Specific repositories for grey literature have been created but the indexing necessary to make documents easily retrievable is incomplete<sup>2, 6</sup>. Generic commercial search engines index only a fraction of these documents which leaves most grey literature inaccessible. PH professionals have expressed interest in informatics tools to help them locate trusted, high-quality PH information on the Web<sup>7</sup>.

Over the last 10 years the Center for Natural Language Processing (CNLP) has sought to address the problem of providing access to a larger number and greater variety of PH grey literature documents using two major approaches. Firstly, we used principles of natural language processing, text mining and user centered design to create document surrogates. Drawing on a model of grey literature documents we extracted key elements of each document to create documents summaries<sup>8</sup>. Secondly we developed a search engine designed to locate quality documents from known PH sites. Domain specific document retrieval is known to outperform generic search engines due to more complete indexing and a focus on high quality relevant resources<sup>9</sup>. A prototype domain specific Web-based search engine, *Public Health Information Search (PHIS)*, was developed to provide the PH workforce with improved access to high quality, highly relevant grey literature reports. The documents were processed using NLP metadata generation techniques to produce a richer search index and to create document summaries based on a PH intervention report model<sup>9</sup>. In this paper we describe the use of a two stage field evaluation to test and inform the iterative development of this Web based PH search engine.

## Methods

### System Development

The PHIS v1 system was developed using open source tools to collect and search documents dealing with chronic disease issues. A commercial search engine was used to search the Web for pages containing any of a list of about 50 chronic disease terms, including pages from about 150 known national, state, and local PH websites. A Web crawl was then conducted starting at these pages, visiting links to the same set of known PH websites, and collecting documents containing any of the chronic disease terms.

The CNLP TextTagger<sup>10</sup> text processing engine was used to generate model summaries for all documents in the collection using a set of rules specialized for PH documents. These rules were derived from previous research where a set of common text elements were extracted from a PH document corpus using NLP techniques. The relative importance of these elements was evaluated through a survey of experts<sup>8</sup>. Elements in the model include document title, problem description, intervention description, results, target population, geographic location, and publication date etc. Model summaries are used for document indexing and to provide an easily understood summary of document content. An example model summary can be seen in Figure 1. TextTagger also identified key words and phrases in the documents. The Apache Lucene<sup>11</sup> search engine was used to index the document collection using TextTagger-identified words and phrases from the documents, as well as words from the TextTagger-generated model summaries.

[ObesityBurdenWebversion.pmd](#)  
is of utmost importance. Unfortunately, **obesity** and weight problems among children and adults have... Initiative to address our state's **obesity** epidemic and to encourage healthy lifestyles for Missourians of all ages. In this report you will find strategies for preventing **obesity** and programs to increase  
<http://www.dhss.mo.gov/Obesity/ObesityBurdenWebversion.pdf>  
[ Model Summaries| Cached ] -- Type: application/pdf

<b>creator</b>	
<b>date</b>	
<b>dates-of-data</b>	1988-2010
<b>description</b>	<ul style="list-style-type: none"><li>• This report provides detailed information about obesity in Missouri .</li><li>• This report addresses some of the most important issues regarding obesity : ?</li><li>• The National Heart , Lung and Blood Institute developed this standard definition of overweight and obesity in 1998 based on an extensive evidenced-based review of morbidity and mortality related to weight .</li><li>• This report was developed under the guidance of the Missouri Department of Health and Senior Services by the Missouri Council on the Prevention and Management of Overweight and Obesity and hundreds of Missourians who provided input on a number of actions that can lead to a healthier Missouri .</li></ul>
<b>disease-or-condition</b>	Obesity
<b>doc-type</b>	Report
<b>geo-location</b>	<ul style="list-style-type: none"><li>• National - United States</li><li>• Missouri</li></ul>
<b>hosted-by</b>	State of Missouri Department of Health and Senior Services
<b>issue</b>	<ul style="list-style-type: none"><li>• Unfortunately , obesity and weight problems among children and adults have increased at astounding rates over the past decade , resulting in health problems that have adversely affected the quality of life for thousands of people and cost the state more than one and a half billion dollars annually .</li><li>• Obesity is a leading cause of preventable death in the United States .</li><li>• Between 1988-2002 , the prevalence of physician-diagnosed diabetes in Missouri adults increased from 4.9 percent to 7.3 percent .</li><li>• Between 1990 and 2002 , the prevalence of obesity among adults in the state increased from 11.9 percent to 23.2 percent ( Figure 1 ) .</li><li>• The prevalence of obesity in Missouri is higher among the following groups of adults : ?</li></ul>
<b>organization</b>	
<b>population</b>	General Population

Figure 1. An example of a PHIS search result with the model summary shown

Users' queries were simultaneously submitted to Google and processed using the Lucene engine. The two ranked result sets were presented side-by-side to the user as shown in Figure 2. The PHIS v1 system searched the PH document collection using TextTagger-identified words and phrases from users' queries, favoring words that were contained in the model summaries.

## PUBLIC HEALTH INFO SEARCH

Query:

Hide Google Results

<p><b>PHIS</b> Results: 1 - 10 of 63 [Results from Index: 300KwithNLM-small-crawl]</p> <ol style="list-style-type: none"><li><b>1. ObesityBurdenWebversion.pmd</b> Missouri Department of Health and Senior Services 2005 Missouri's Nutrition and Physical Activity Plan Preventing Obesity and Other Chronic Diseases Healthy Missourians Initiative Preventing Obesity and Other Chronic Diseases Missouri's Nutrition an... <a href="http://www.dhss.mo.gov/Obesity/ObesityBurdenWebversion.pdf">http://www.dhss.mo.gov/Obesity/ObesityBurdenWebversion.pdf</a> [ Model Summaries   Cached ] -- Type: application/pdf</li><li><b>2. Research Directions in Childhood Obesity Prevention and Trea ...</b> Background and Objectives Childhood overweight and <b>obesity</b> have reached epidemic... overweight and <b>obesity</b> add to the burden of health-care costs. The annual U.S.... were higher than those that were not obesity-associated (e.g., discharges with <b>obesity</b> as a <a href="http://www.nhlbi.nih.gov/meetings/workshops/child-obesity/in...">http://www.nhlbi.nih.gov/meetings/workshops/child-obesity/in...</a> [ Model Summaries   Cached ]</li><li><b>3. Full Report of the Working Group on Competencies for Overwei ...</b> References A. Background and Purpose Overweight and <b>obesity</b> have reached epidemic proportion... Although the prevalence of overweight or <b>obesity</b> among adults at least 20 years of age is lower in... to tackle the <b>obesity</b> epidemic from multiple fronts, including training medical students, resident <a href="http://www.nhlbi.nih.gov/meetings/workshops/overweight/repor...">http://www.nhlbi.nih.gov/meetings/workshops/overweight/repor...</a> [ Model Summaries   Cached ]</li></ol>	<p><b>Google</b> Results: 1 - 10</p> <ol style="list-style-type: none"><li><b>1. Obesity - Wikipedia, the free encyclopedia</b> <b>Obesity</b> is a medical condition in which excess body fat has accumulated to the extent that it may have an adverse effect on health, leading to reduced life ... <a href="http://en.wikipedia.org/wiki/Obesity">http://en.wikipedia.org/wiki/Obesity</a> - Cached</li><li><b>2. Obesity and Overweight: Topics   DNPAO   CDC</b> This site provides a variety of information designed to help people understand that <b>obesity</b> is a serious health issue and the efforts being made to address ... <a href="http://www.cdc.gov/obesity/">http://www.cdc.gov/obesity/</a> - Cached</li><li><b>3. Obesity Facts, Causes, Emotional Aspects and When to Seek Help</b> <b>Obesity</b> is an excess proportion of total body fat. A person is considered obese when his or her weight is 20% or more above normal weight. <a href="http://www.webmd.com/diet/guide/what-is-obesity">http://www.webmd.com/diet/guide/what-is-obesity</a> - Cached</li><li><b>4. Obesity.org</b> The <b>Obesity</b> Society is the leading scientific society dedicated to the study of <b>obesity</b>. Since 1982, The <b>Obesity</b> Society has been committed to encouraging ... <a href="http://www.obesity.org/">http://www.obesity.org/</a> - Cached</li></ol>
--	---

**Figure 2.** The PHIS results page for the query term “obesity”. PHIS collection results are on the left (with model summaries hidden), Google results are on the right.

Midway through the first evaluation, results indicated that all of the subjects were dissatisfied with specific search results: additional query language capabilities (e.g. “AND” and “OR”) were consistently requested and subjects complained that results from information “sparse” websites such as state health departments dominated over preferred results from information “dense” sites such as trusted national organizations (e.g. American Lung Association, CDC). Because of the iterative nature of the development process, the evaluation was halted for one week while query language capability was added and the search catalog was updated with federal agency and national organization websites. After this update the system was still referred to as PHIS v1.

Based on evaluation of the PHIS v1 system, two main areas were identified for improvement in PHIS v2: increasing the effective size of the document collection to be searched and improving the quality of the document rankings produced in response to queries. The basic look and feel of the v1 system was retained.

### *Increased Collection Size*

The PHIS v1 collection was specifically targeted to include documents dealing with chronic disease but the first evaluation showed that PH experts had search interests that were much broader than the collection focus. In order to expand the collection of documents we adopted three strategies: 1) increase the number of websites visited to increase the number of candidate documents considered, 2) relax the chronic disease filter to increase the number of documents accepted from the candidate set for inclusion in the PHIS collection, and 3) draw results from both the PHIS collection and a search of the entire Web.

Increasing the number of documents in the PHIS collection is conceptually straightforward; we simply increased the range of sites visited when collecting documents and adjusted filter settings to favor medical documents rather than only chronic disease documents. As a result, the actual collection size grew from roughly 11,000 in the PHIS v1 collection to 260,000 in the PHIS v2 collection. There is some overlap between the PHIS collection and the collection indexed by any of the Web search engines but a substantial proportion of the PHIS documents are not indexed by the Web search engines.

Drawing documents from a search of the entire Web was accomplished using a programmatic interface provided by a Web search engine (Yahoo!). The results of the Web search were combined with the results from the local PHIS collection to form a single document ranking that was presented to the user. The document summaries produced for the Web documents were less complete than the summaries produced for the local PHIS collection since the NLP processing required to generate the full summary for previously unseen documents would have introduced a response time delay.

### Improved Ranking

The ranking used in the PHIS v1 system was based on the normal Lucene ranking algorithm. Early experiments comparing the Lucene ranking with more advanced ranking functions based on language models<sup>12, 13</sup> using the open source Lemur search engine<sup>14</sup> showed that significant improvements were possible. These initial tests were conducted with standard TREC test collections<sup>15</sup>.

In order to test ranking performance with the PHIS v2 collection we developed a set of 98 test queries dealing with PH topics together with pooled relevance judgments from Lucene and Lemur searches. These test queries and relevance judgments were used to test the effect of changes made to the baseline rankings.

The PHIS ranking function combines documents from two sources (the PHIS collection and the Yahoo! Web search). For the PHIS collection we have two independent ranking functions: Lemur and Lucene. While our experiments showed that the Lemur ranking was consistently superior we found that combining the two rankings produced a better result than either individual ranking alone. This “boosting” effect is common when combining independent rankings<sup>16, 17</sup>. The final PHIS v2 ranking combines the Lemur, Lucene, and Yahoo! results using a weighted zipper merge<sup>18</sup>.

We initially hoped to integrate UMLS vocabulary resources for indexing, document summary generation, and query expansion. It was not possible to complete this work due to time and budget limitations; this is an important item of future research.

### First Evaluation

The first field evaluation employed the Critical Incident Technique, developed during World War II studies of aviation psychology<sup>19</sup>. The technique involves collecting a set of incidents in which a qualified individual judges the outcome of some activity as either especially effective or ineffective. In this evaluation an incident was considered a single search session. Each search session incident may include only a single search, or multiple search refinements, but is restricted to a single topic during a single contiguous time period. The resulting collection of incidents is analyzed to identify critical requirements for the activity. This empirical data provides a more specific and valid basis for evaluation than requirements based on expert opinion. This study design was chosen because of its effective use in evaluating information retrieval systems<sup>20</sup>. The characteristics of both the first and second evaluations are shown in Table 1.

**Table 1.** Characteristics of the first and second evaluations and each version of PHIS.

Evaluation	Num. of Subjects	Evaluation period	PHIS version	Collection size	Content focus	Num. of critical incident searches evaluated	Num. of other searches evaluated
1	10	4 weeks	PHIS v1	11,000	Chronic diseases	28	--
2	6	4 weeks	PHIS v2	260,000	No focus	42	124

A convenience sample of ten PH practitioners was recruited from two large full service county health departments and an urban community health center. Selection was based on the criterion that the subjects conducted Web searches for PH information related to chronic disease topics as part of their routine work. Cash incentives were offered. The sample consisted of 2 biostatisticians, 2 data analysts, 1 epidemiologist, 1 PH nurse, 2 health educators, 1 community health worker, and 1 program director. The research team traveled to the county health departments to train eight subjects on use of the system. The community health worker and program director were trained remotely. One subject dropped out of the evaluation after week 3 citing time constraints.

The evaluation took place over a period of 4 weeks. Each subject was asked to use PHIS v1 during their normal work, particularly work oriented toward chronic diseases. Subjects were instructed to take notes during the week on their experience performing a search (i.e. an “incident”) in which PHIS was particularly effective or ineffective. Although effectiveness is an inherently subjective measure that is purposefully not defined in the critical incident technique, subjects generally determined a search incident to be effective if the PHIS results both addressed the information need and were as good or better than the Google results. Subjects were instructed to report on either an effective or ineffective report each week.

At a scheduled time a researcher conducted a retrospective phone interview with each subject. Potential reporting bias was addressed by using an interview guide adapted from Lindberg's seminal evaluation of MEDLINE and by maintaining a consistent semi-structured interviewing technique<sup>20</sup>. The interview guide is shown in Table 2. All but 2 of the interviews were conducted by a single graduate student and all interviews took place under the guidance of a senior researcher.

**Table 2.** Interview guide adapted from Lindberg's MEDLINE evaluation.

1	What situation led you to do this search?
2	What specific information were you seeking?
3	What were the search terms used and in what order?
4	Did you use the model summaries? Were they helpful? Did they save you time?
5	What information did you obtain?
6	Why were search results satisfactory/unsatisfactory?
7	Did Google satisfy your information query?
8	How was the information helpful in your decision making?
9	Did PHIS help focus your information needs?
10	How did having the information impact your work?
11	Did another information source help you find the information you needed?
12	Was it worth it to use PHIS versus what you normally would have done?
13	Any final thoughts?

During weeks 1 and 2 of the evaluation several systematic but readily addressed problems with the searching protocols emerged. Because this was an iterative development process, the research team decided to take a pragmatic approach by suspending the evaluation for one week to allow the developers to address the problems.

### *Second Evaluation*

The second field evaluation took place approximately 16 months after the first. The intervening time was used to develop PHIS v2. A convenience sample of six PH practitioners was recruited from three large full service county health departments. The difference in sample size between the evaluations was due to time constraints. Selection was based on the criterion that the subjects conducted Web searches for PH information as part of their routine work however the chronic disease requirement was dropped. Cash incentives were offered. The sample consisted of 1 epidemiologist, 2 PH nurses, 2 health educators, and 1 emergency response coordinator. Three of the subjects participated in the first evaluation and three were new. All subjects were trained remotely.

The evaluation methods and interview guide used were identical to the first evaluation with the following two exceptions: 1) Subjects were asked to report two incidents each week; one in which PHIS v2 was particularly effective and one in which it was particularly ineffective, and 2) Subjects were asked to keep a simple tally of PHIS v2 effective versus ineffective for all PH related searches during the week, not just the critical incidents. The effectiveness of Google results for these searches was not recorded. The changes to the evaluation methods were made to accommodate the smaller sample size by collecting more observations.

Data from both evaluations were analyzed using the constant comparison/grounded theory technique to assess why the information was needed, the effective rate of PHIS and the model summaries in comparison to Google results, and the factors that impacted the usefulness of PHIS and the model summaries.

## **Results**

### *Search Motivation*

The motivation for the search incidents in both evaluations fell into 4 themes. The first theme was preparation of official documents such as creating a community health assessment or emergency response plan. The second theme was preparation for publications, presentations, or websites. Examples include finding tips for a website health campaign and searching for background information for a talk on body image. The third theme was searching for information on a specific disease or condition related to projects or research. Examples include determining the origin and causes of MRSA and studying the different types of salmonella lab tests. The final theme was fielding questions from the public such as proper caloric intake for a person with diabetes. The counts and effective rates for each of these themes are shown in Table 3.

**Table 3.** Counts and effective rates for the four search motivation themes.

Search Motivation	First Evaluation		Second Evaluation	
	Count	Effective rate	Count	Effective rate
Preparation of Official Documents	9	11%	6	83%
Preparation for publications, presentations, or websites	7	14%	15	60%
Seeking information on specific diseases/conditions	10	20%	21	33%
Fielding questions from the public	2	50%	0	--

### *System Effectiveness*

Of the 28 search incidents collected during the first evaluation, PHIS v1 was effective in 5, giving an effective rate of 18%. Google was effective in 14 of 28 or 50%. When 14 searches that were not specific to the document collection focus of chronic diseases are removed from the calculations the effective rate of PHIS v1 is 28% and the effective rate of Google is 64%. Out of 166 total searches in the second evaluation, PHIS v2 was effective in 91 for an effective rate of 55%. Critical incident reports were collected on 42 of those searches, 21 effective and 21 ineffective. Of the 21 ineffective search incidents, Google was effective for 17.

The model summaries were used in 21 of the search incidents during the first evaluation and were found to be useful in 81% of those searches. In the second evaluation the model summaries were used in 32 of the search incidents and were found to be useful in 94% of those searches. In both evaluations subjects indicated that the model summaries were more useful than Google page excerpts. Subjects appreciated how data was parsed into different fields and found the date, geo-location, and description fields particularly useful. All subjects that used the model summaries stated that they saved time, specifically because they could avoid opening non-relevant PDF files.

During the first evaluation, subjects indicated three primary issues that prevented PHIS v1 from being more effective. The first was lack of geographic focus. Thirteen (46%) of the searches were geographically specific and only one of those was effective. The ineffectiveness of these searches was not due to validity of the topic, but inappropriateness of the geographic localization. An example of this type of search would be “breast cancer statistics” which was intended to retrieve national level data but instead retrieved breast cancer statistics for states such as Utah, Michigan, and others.

The second issue was relevance. Eleven (39%) of the searches returned results that were considered irrelevant to the topic of the search. An example is a search query “obesity premature pregnancy” which returned results on obesity and results concerning premature pregnancy but no results about co-morbidity. The third issue was trust. One of the key aspects on which the subjects evaluated usefulness of results was trust in the source of the information. This issue was part of the impetus for suspending the evaluation of the v1 system for reprogramming. Many subjects were looking for information from national sites. They indicated their reason for this was that they trusted information from sites such as the CDC, the NIH, and national associations such as the National Cancer Institute. The grey literature documents in search results often came from the state level and were not as trusted. Most of the search incidents where trust was an issue were searches for specific disease/condition information.

During the second evaluation subjects were more positive about PHIS v2. They indicated that PHIS was most useful when it returned resources and documents that normally would not be found in mainstream search engines. These results served to complement Google results and included non-indexed reports, white papers, and more technical documents such as templates. Subjects also indicated trust was no longer an issue as they found an adequate mixture of both obscure documents and national mainstream sites in the PHIS results. Relevance had also increased as PHIS v2 returned no results that were completely off topic versus 39% in PHIS v1. Improvement in geographic focus was difficult to assess. It was not mentioned as a specific problem, however only 17% of searches in the second evaluation included a geographic component compared to 46% in the first evaluation.

Subjects indicated one primary issue with PHIS v2 that impacted effectiveness. Subjects reported that PHIS v2 results were sensitive to the number and combination of search terms in comparison to Google results. The quality of PHIS v2 results was adversely affected both by too many search terms and by terms that were too generic. For example, in a search for “heart disease symptoms diagnosis” PHIS v2 would return a mix of results on other diseases because they contained the terms “disease” and “symptoms” and would ignore the term “heart”. Subjects found that PHIS v2 was more effective if fewer, more specific terms were used. Google results did not exhibit this problem.

All subjects in both evaluations stated they would continue to use PHIS if it were offered in conjunction with Google. Subjects reacted favorably to the side by side screen display of PHIS with Google. It did not impede their normal search experience and/or workflow and allowed them access to extra results. Subjects' feedback during the final week included feature requests such as limiting results by geographic level (national, state, local), and by theme (e.g., policy and legislative, statistics, disease information).

## **Conclusion**

A series of two field evaluations helped with the iterative design of PHIS. By expanding the document collection size from 11,000 to 260,000 and improving the ranking algorithm PHIS search effectiveness was improved from 28% to 55%. Both the relevance and trust issues were successfully addressed. The results suggest that online searching is a primary and effective means of information gathering in PH that can be improved with a grey literature search engine. Although this study did not evaluate the significance of grey literature to PH information seeking, the high effective rate suggests grey literature can be useful.

By dividing the effective rate between the search motivation themes in Table 3, we can see that PHIS is more effective at searching for information related to creation and preparation of official documents and for preparing publications and presentations. The types of information being sought in these instances include document templates, example reports, department white papers and appendices. PHIS appears to be less effective than Google at returning information on specific diseases and conditions. This suggests that PHIS is better at indexing government and other agency documents that are uploaded to a variety of departmental websites, and not as good at indexing disease information from trusted national websites.

While the subjects had increased trust in the results of PHIS v2 because national and federal websites such as the CDC, MedlinePlus, and the National Cancer society were included, these are also the type of sites already well indexed by Google. Paradoxically, the subjects found the most useful PHIS results were the grey literature results that Google could not find such as document templates and appendices. This apparent contradiction may be a fault of evaluation design. Both evaluations were designed to compare the effectiveness of PHIS and Google and not to assess how they worked together in a paired manner. This may have led the subjects to conclude that PHIS was meant to replace Google and therefore should index everything they found valuable in Google results as well as grey literature. This limitation should be addressed in future evaluations. The issue of what makes information trustworthy to a PH practitioner was outside the scope of this research, but is an important area for future work. A limitation which we could not control for was the variability in the type of information being sought between the evaluations which should be considered when interpreting the results.

The auto-generation of model summaries of documents, especially of those documents that are not in HTML and may be slow to open, was favored by the subjects and would be useful for PH practitioners. Unfortunately, full model summaries were only generated on the PHIS document collection. Yahoo! results were limited to incomplete model summaries due to technical constraints. Faster NLP and architectural modifications in the future would allow complete summaries to be generated for all documents in a production system. Future work should also include continued improvement of the algorithm for using search terms or phrases. This might include decreasing the weight of more general terms in favor of more specific terms.

By using PHIS side by side with Google or another commercial search engine, the quantity of results presented to the user is increased without affecting user efficiency or workflow. Because PHIS indexes items and ranks them specifically for PH practice, the quality of search results is improved and by using the model summaries to more quickly find relevant documents without having to open them, PHIS can save PH practitioners valuable time. PHIS has the potential to help improve the quantity, quality, and speed of information retrieval for PH practitioners, especially when searching for information that is found in grey literature.

## **Acknowledgements**

This research was supported by the National Library of Medicine G08 LM008983 "Improving PH Grey Literature Access for the Public Health Workforce." The research team wishes to express gratitude to PH practitioners in New York's Hudson Valley for their participation in the evaluation. Keeling is a National Library of Medicine pre-doctoral trainee supported by NLM grant T15-LM007079.

## References

1. Revere D, Turner A, Madhavan A, *et al.* Understanding the information needs of PH practitioners: A literature review to inform design of an interactive digital knowledge management system. *JBI*. 2007; 40(4):410-421.
2. LaPelle N, Luckmann R, Simpson E, Martin E. Identifying strategies to improve access to credible and relevant information for PH professionals: a qualitative study. *BMC PH*. 2006; 6(89).
3. Lasker R. Challenges to accessing useful information in health policy and PH: an introduction to a national forum held at the New York Academy of Medicine. *J Urban Health*. 1998; 75: 779-84.
4. GL '99 Conference Program. Fourth International Conference on Grey Literature: New frontiers in grey literature. Washington, D.C. GreyNet, Grey Literature Network Service. 4-5 Oct. 1999.
5. Tomaiuolo NG, Packer JG. Preprint servers: pushing the envelope of electronic scholarly publishing. *Searcher*. 2000; 8(9).
6. Myohanen L, Taylor E, Keith L. Accessing grey literature in PH: New York Academy of Medicine's grey literature report. *Publishing Research Quarterly*. 2005; 21(11): 44-52.
7. Turner A.M, Petrochilos D, Nelson DE, Allen E, Liddy ED. Access and use of the internet for health information seeking: a survey of local public health professionals in the northwest. *JPHMP*. 2009; 15(1): 67-9.
8. Turner A, Liddy E, Bradley J, Wheatley J. Modeling PH interventions for improved access to the gray literature. *J Med Libr Assoc*. 2005; 93(4): 487-94.
9. Liddy E. Question answering in contexts. AQUAINT Annual Meeting. Washington DC. Dec 2-5, 2003.
10. Ingersoll G, Yilmazel O, Liddy E. Finding questions to your answers. *IEEE Data Engineering Workshop*. 2007: 755-759.
11. Deng-peng Z. Lucene search engine. *Computer Engineering*. 2007; 18.
12. Croft WB, Lafferty J. *Language modeling for information retrieval*. Kluwer, 2003.
13. Zhai C. *Statistical language models for information retrieval*. Morgan & Claypool, 2008.
14. Allan J, Callan J, Collins-Thompson K, *et al.* The Lemur toolkit for language modeling and information retrieval. <http://www.cs.cmu.edu/lemur/>, 2003.
15. Voorhees M, Harman D. *TREC: Experiment and evaluation in information retrieval*. MIT Press. 2005.
16. Savoy J, Le Calve A, Vrajitoru D. Data fusion and collection fusion. *Proceedings of TREC-5, NIST Publication 500-238*, 1997; 489-502.
17. Callan J. Distributed information retrieval. In: *Advances in Information Retrieval*. 2002; 7:127-150.
18. Towell G, Voorhees E, Gupta N, Jouhson-Baird B. Learning collection fusion strategies for information retrieval. *Proceedings of the 12<sup>th</sup> Annual Machine Learning Conference*. 1995.
19. Flanagan JC. The critical incident technique. *Psychol Bull*. 1954; 51: 327-58.
20. Wilson SR, Starr-Schneidkraut N, Cooper MD. *Use of the critical incident technique to evaluate the impact of MEDLINE*. Palo Alto, CA: American Institute for Research, 1989.