10-30-2016

Ontology-Driven Search and Triage: Design of a Web-Based Visual Interface for MEDLINE

Jonathan Demelo
Insight Lab, Department of Computer Science, Western University, London, ON, Canada

Paul Parsons
Purdue Polytechnic Institute, Department of Computer Graphics Technology, Purdue University, parsonsp@purdue.edu

Kamran Sedig
Insight Lab, Department of Computer Science, Western University, London, ON, Canada

Original Paper

Ontology-Driven Search and Triage: Design of a Web-Based Visual Interface for MEDLINE

Jonathan Demelo[1*], BSc; Paul Parsons[2*], PhD; Kamran Sedig[1*], PhD

[1]Insight Lab, Department of Computer Science, Western University, London, ON, Canada

[2]Purdue Polytechnic Institute, Department of Computer Graphics Technology, Purdue University, West Lafayette, IN, United States

[*]all authors contributed equally

Corresponding Author:
Paul Parsons, PhD
Purdue Polytechnic Institute
Department of Computer Graphics Technology
Purdue University
Knoy Hall
401 N Grant St
West Lafayette, IN, 47907
United States
Phone: 1 765 494 0511
Fax: 1 765 494 9267
Email: parsonsp@purdue.edu

Abstract

Background: Diverse users need to search health and medical literature to satisfy open-ended goals such as making evidence-based decisions and updating their knowledge. However, doing so is challenging due to at least two major difficulties: (1) articulating information needs using accurate vocabulary and (2) dealing with large document sets returned from searches. Common search interfaces such as PubMed do not provide adequate support for exploratory search tasks.

Objective: Our objective was to improve support for exploratory search tasks by combining two strategies in the design of an interactive visual interface by (1) using a formal ontology to help users build domain-specific knowledge and vocabulary and (2) providing multi-stage triaging support to help mitigate the information overload problem.

Methods: We developed a Web-based tool, Ontology-Driven Visual Search and Triage Interface for MEDLINE (OVERT-MED), to test our design ideas. We implemented a custom searchable index of MEDLINE, which comprises approximately 25 million document citations. We chose a popular biomedical ontology, the Human Phenotype Ontology (HPO), to test our solution to the vocabulary problem. We implemented multistage triaging support in OVERT-MED, with the aid of interactive visualization techniques, to help users deal with large document sets returned from searches.

Results: Formative evaluation suggests that the design features in OVERT-MED are helpful in addressing the two major difficulties described above. Using a formal ontology seems to help users articulate their information needs with more accurate vocabulary. In addition, multistage triaging combined with interactive visualizations shows promise in mitigating the information overload problem.

Conclusions: Our strategies appear to be valuable in addressing the two major problems in exploratory search. Although we tested OVERT-MED with a particular ontology and document collection, we anticipate that our strategies can be transferred successfully to other contexts.

(JMIR Med Inform 2017;5(1):e4) doi:10.2196/medinform.6918

KEYWORDS

MEDLINE; user-computer interface; information storage and retrieval; medical informatics; PubMed

Introduction

Overview and Significance

Seeking information within the published medical literature is important in many domains and contexts [1,2]. Diverse users need to search the literature including physicians [3], medical students [4], cytogeneticists [5], and patients and their relatives [6]. Searches can be roughly categorized into 2 types: *lookup* and *exploratory* [7]. Lookup searches are closed-ended, having precise results and little need for examining and comparing result sets. Exploratory searches, however, are open-ended, having imprecise results and often requiring significant time and effort to work with result sets in order to satisfy the original information need. Examples of exploratory searches with open-ended goals include making evidence-based decisions and updating knowledge to stay abreast of current research findings [2,8]. Although significant progress has been made in supporting lookup searches, exploratory searches are still not well supported, and open-ended search goals are often quite difficult to achieve [2,9,10]. Common barriers to finding relevant medical information include the time it takes to perform searches [3,11], the increasing scope of topical coverage [2], and the information overload that arises from dealing with large result sets [2,3,11-13].

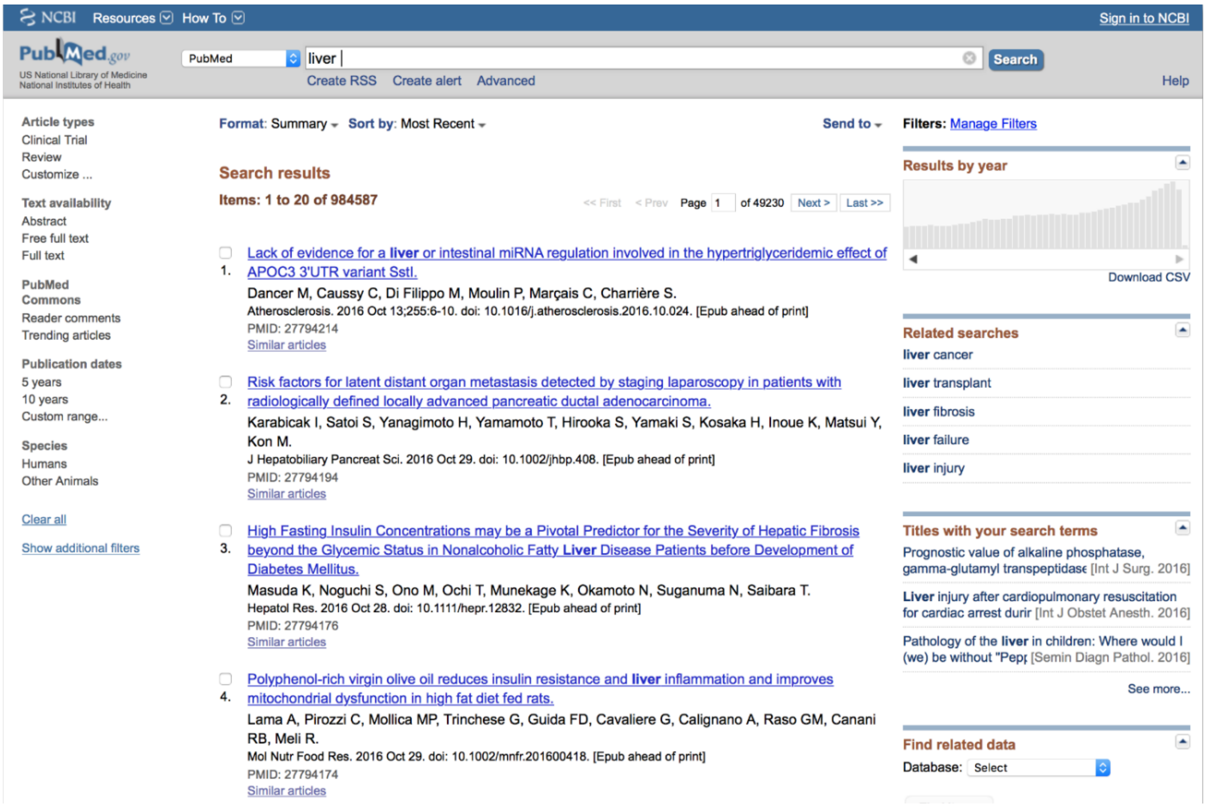
One of the most popular collections of published medical literature is MEDLINE, which comprises more than 25 million documents and is growing every year. The most common means of searching MEDLINE is PubMed, a free search engine and Web interface [14]. Although the search capabilities in PubMed have improved in recent years, there can still be a considerable burden on users when seeking information in the context of exploratory search, due to at least two major problems: (1) the difficulty in articulating information needs using accurate vocabulary and (2) the large number of documents that can be returned from searches. Many users do not have the proper vocabulary to construct effective queries [15,16], which is especially true in medical and health contexts [17-20]. When uncontrolled vocabularies are used, there is no guarantee that concepts are expressed with the same terms in different contexts [13,21]. For instance, if an article contains the term *eye hamartoma*, and a user searches for the vaguer term *eye growth*, there may not be a close match. Thus, without proper terminological knowledge, effective searching can be quite difficult. Adding to the difficulty of searching effectively is the large number of documents that can be returned, which leads to information overload problem [9,22,23]. Dogan et al [2] note that at least one-third of PubMed searches return 100 or more documents. In our own testing, searches for common terms (eg, “breast cancer” or “brain tumor”) returned many thousands of documents.

Interfaces to most search engines, including PubMed, use simple text boxes into which users enter query terms. This interface style does not assist users in articulating their information needs [24] and works well only for lookup search tasks [25,26]. For example, if a user is interested in finding information about “liver,” but is not sure what terms are relevant in articulating a query, he or she must simply enter “liver” into the search box. As the query is vague, a very large set of documents is returned—almost one million documents spanning over 4900 pages when using PubMed (Figure 1).

Multiple strategies have been employed to help support query formation in exploratory search contexts by replacing the standard text box, including faceted search [27], visualization widgets [28], query previews [29], and hierarchical presentation of expansion terms [30]. The common theme among these strategies is that meaningful information is extracted from the document collection and then represented in a manner that can help the searcher recognize terms that will more accurately describe the information they are seeking. Such strategies promote recognition over recall, not relying on users having to know and retrieve correct vocabulary from memory [24].

We present Ontology-Driven Visual Search and Triage Interface for MEDLINE (OVERT-MED), a Web-based visualization tool that addresses two major difficulties in searching large document collections: (1) the difficulty in articulating information needs with useful vocabulary and (2) the difficulty in dealing with large search result sets. To address the first difficulty, we propose the idea of using a formal ontology to help users build domain-specific knowledge and vocabulary. To test this, we have implemented a searchable index of the Human Phenotype Ontology (HPO) that provides users with suggestion terms that are related to their information needs. To address the second difficulty, OVERT-MED supports multistage interactive triaging of search results using interactive visualization techniques. We use a custom-built index of MEDLINE, which comprises approximately 25 million documents, as our searchable collection of medical literature. Although OVERT-MED has been initially developed for use with a particular ontology and document collection, we expect that our design ideas will transfer to other contexts. The following subsections provide background information and discuss related work.

Figure 1. A screenshot of PubMed showing results from searching for “liver.”



Search results

Items: 1 to 20 of 984587

□ 1. Lack of evidence for a **liver** or intestinal miRNA regulation involved in the hypertriglyceridemic effect of APOC3 3'UTR variant SstI.
Dancer M, Caussy C, Di Filippo M, Moulin P, Marçais C, Charrière S.
Atherosclerosis. 2016 Oct 13;255:6-10. doi: 10.1016/j.atherosclerosis.2016.10.024. [Epub ahead of print]
PMID: 27794214
Similar articles

□ 2. Risk factors for latent distant organ metastasis detected by staging laparoscopy in patients with radiologically defined locally advanced pancreatic ductal adenocarcinoma.
Karabicak I, Satoi S, Yanagimoto H, Yamamoto T, Hirooka S, Yamaki S, Kosaka H, Inoue K, Matsui Y, Kon M.
J Hepatobiliary Pancreat Sci. 2016 Oct 29. doi: 10.1002/jhbp.408. [Epub ahead of print]
PMID: 27794194
Similar articles

□ 3. High Fasting Insulin Concentrations may be a Pivotal Predictor for the Severity of Hepatic Fibrosis beyond the Glycemic Status in Nonalcoholic Fatty **Liver** Disease Patients before Development of Diabetes Mellitus.
Masuda K, Noguchi S, Ono M, Ochi T, Munekage K, Okamoto N, Suganuma N, Saibara T.
Hepatol Res. 2016 Oct 28. doi: 10.1111/hepr.12832. [Epub ahead of print]
PMID: 27794176
Similar articles

□ 4. Polyphenol-rich virgin olive oil reduces insulin resistance and **liver** inflammation and improves mitochondrial dysfunction in high fat diet fed rats.
Lama A, Pirozzi C, Mollica MP, Trinchese G, Guida FD, Cavaliere G, Calignano A, Raso GM, Canani RB, Meli R.
Mol Nutr Food Res. 2016 Oct 29. doi: 10.1002/mnfr.201600418. [Epub ahead of print]
PMID: 27794174
Similar articles

Titles with your search terms
Prognostic value of alkaline phosphatase, gamma-glutamyl transpeptidas‹ [Int J Surg. 2016]
Liver injury after cardiopulmonary resuscitation for cardiac arrest durir [Int J Obstet Anesth. 2016]
Pathology of the **liver** in children: Where would I (we) be without "Pep[[Semin Diagn Pathol. 2016]
See more...

Ontologies

One way to meaningfully extract and model information from a domain is to construct an ontology [31,32]. An ontology represents concepts and their relationships using a standard vocabulary [32]. Ontologies serve many practical functions, including clarifying the structure of knowledge within a domain, providing a common vocabulary, enabling computational analysis, and supporting knowledge sharing [31-33]. Ontologies often capture concepts within a domain at multiple levels of abstraction. For instance, an anatomy ontology may have a concept *body*, a sub-concept *face*, a further sub-concept *nose*, and so on. The concepts in an ontology can be represented using many different structures, including trees and different types of graphs.

The ontology we are using, HPO, has been curated by domain experts in an attempt to capture all phenotypic abnormalities that are commonly encountered in human monogenic disease [34]. In our previous work with genomics researchers, we learned of the importance of HPO in their workflow, including in activities involving literature search [5]. HPO is widely used in the biomedical field, is regularly updated, and has a high level of quality control. It is also available for download in the popular Open Biomedical Ontologies (OBO) and Web Ontology Language (OWL) formats. For these reasons, we believe HPO is ideal for testing our proposal of using ontologies to address the vocabulary problem. It should be noted that we are not suggesting HPO is better than other ontologies or that it should be used in all contexts. HPO is only one of the many ontologies that could be used to support exploratory search, and search systems should make use of whichever ontologies are most appropriate for given contexts.

Document Triage

Triaging is an activity that involves determining the relevance of documents to an information need [35]. Triaging activities are often time-constrained and require quick assessment of relevance with incomplete knowledge. For example, a search may return hundreds or thousands of potentially relevant documents. As it is not feasible to read each one in detail, users must sort through the documents and quickly assess their relevance based on incomplete knowledge of their contents. Research suggests that triaging takes place in 3 successive stages: (1) the “multiple document” stage, where initial relevance judgments are made to select documents from a set without careful examination; (2) the “individual document” stage, where individual documents are examined in more detail and categorized (eg, kept or rejected); and (3) the “further reading” stage, where a small set of documents are read in depth to extract relevant information and satisfy the original information need [36]. In addition, research shows that triaging often occurs in a cyclical and iterative fashion, where the above stages are revisited multiple times [37].

Search Result Visualization

Most search interfaces present results in a traditional list-based manner, where documents are ranked and textually represented using a title and various metadata. While not a problem for simple lookup search tasks, traditional list-based representations are not effective in supporting exploratory search tasks, which are typically open-ended and involve complex information needs [38]. Although lists are familiar and simple, studies show that users rarely examine lists fully or carefully [39] and seldom venture past the first few pages of results [40]. Scanning through long lists can be tedious and cognitively demanding.

Visualizations of search results can overcome some of the problems associated with textual list-based representations by shifting cognitive burden onto the perceptual system. For instance, whereas visualizations can be scanned freely by the eyes, text must be scanned sequentially, requiring more time and cognitive effort to detect patterns and relationships [41,42]. In addition, visualizations can encode a significant amount of information within a small space, removing the need to navigate multiple pages to view search results. Previous work has demonstrated the utility of visualizations in document search, exploration, and analysis [43,44].

Related Work

Some researchers have recognized the value of using ontologies to better support search activities (eg, [13,45]). The central focus of this research is term extraction and mapping, which is done using text mining and natural language processing techniques. In this body of work, ontologies are used to improve search performance computationally without involving users. The fundamental difference compared with our work is that we use ontologies to help users develop knowledge and domain-specific vocabulary—that is, the focus is on the user rather than on algorithms and other computational processes. Our approach is important in contexts where users have valuable knowledge and context-specific goals that cannot be replaced by computation—in other words, users need to be kept "in the loop."

Other researchers have focused on developing interfaces to MEDLINE as alternatives to PubMed. For example, Wei et al have developed PubTator, a PubMed replacement interface that uses multiple text mining algorithms to improve search results [46]. PubTator also offers some support for document triaging. Whereas PubTator appears interesting and useful, it relies on queries being input into the standard text box, and it presents results in a typical list-based fashion. Thus, it is not aimed at addressing either of the two problems we are attempting to address with OVERT-MED—that is, the vocabulary problem and the information overload problem. Other alternative interfaces that offer interesting features but do not address either of the two problems include SLIM [47] and HubMed [48]. An alternative interface that potentially provides support in addressing the first problem is iPubMed [49], which provides fuzzy matches to search results. An alternative interface that may provide support in addressing the second problem is refMED [50], which provides minimal triaging support through relevance ranking. A for-profit private tool, Quertle, appears to use visualizations to mitigate the information overload problem, although very few details are publicly available. Lu [51] provides a detailed survey that includes many other alternative interfaces to MEDLINE, although none are aimed at solving either of the two problems that we are addressing here.

In summary, no extant research explores the combination of (1) ontologies to help build domain-specific knowledge and vocabulary when users need to be kept "in the loop" and (2) triaging support using interactive visualizations to help mitigate the information overload problem. The following sections provide details about our approach to addressing these issues.

Methods

Overview

We developed OVERT-MED to test our proposed solutions to the two problems described hereinbefore. To anchor our research in a specific context, we chose MEDLINE as our document collection. MEDLINE offers an interesting testbed because of its popularity and size. We developed a custom index of MEDLINE so that it can be queried from the front end of OVERT-MED. We have also indexed HPO to help users build knowledge and domain-specific vocabulary.

Indexing of MEDLINE and HPO

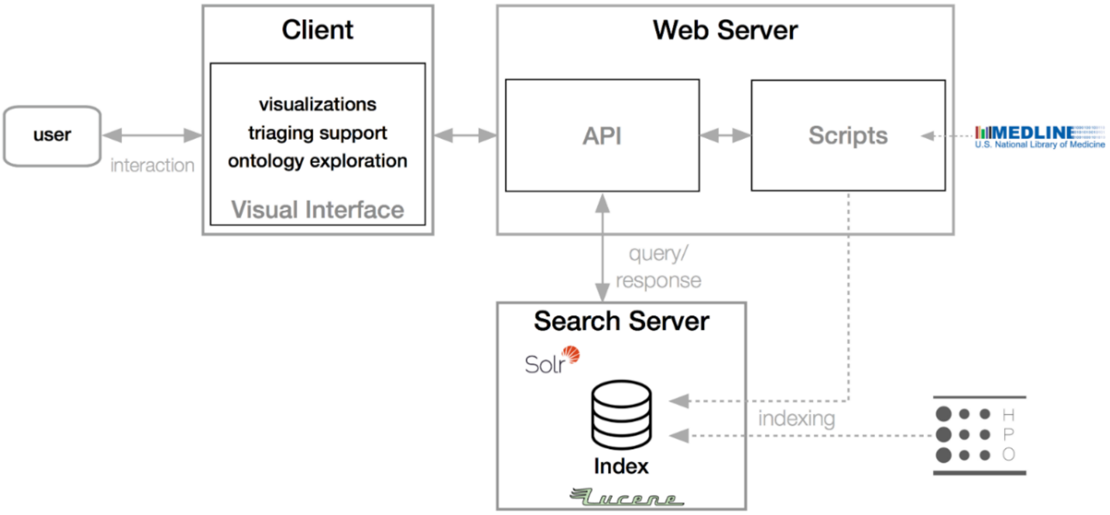
We downloaded the entire MEDLINE database, which has been made freely available by the National Library of Medicine (NLM) for research purposes. The MEDLINE database consists of article "citations," which are essentially article metadata, including authors, journal title, Medical Subject Heading (MeSH) keywords, publication date, and other fields. Also included in each citation is the abstract text. We developed a custom index using the open-source Apache Solr and Lucene projects. Lucene supports full-text indexing and search functionality, and Solr is a search platform that runs on the Lucene index. To rank documents, Lucene uses the well-known term frequency-inverse document frequency (tf-idf) scheme [52]. Lucene also ranks results based on an internal similarity measure that generates a vector space model (VSM) score [53], using index terms as dimensions and tf-idf values as weights. We have described our indexing strategy in greater detail earlier [5].

HPO is a formal ontology of human phenotypic abnormalities found in human disease [34]. Each entry in HPO describes a phenotypic abnormality such as melanoma or hepatoblastoma. HPO is under active development and currently contains more than 11,000 terms. We have also indexed HPO in our Lucene index. HPO contains multiple fields for each phenotype in the ontology, including name, definition, id, synonyms, and commentary from domain experts. We index all fields to provide robust vocabulary suggestions—when a user enters a term, all fields in the index are examined, which provides much more useful information than would result from looking for only exact matches on the phenotype name. This is described using an example in greater detail in the following.

Development and Architecture

We developed OVERT-MED as a Web-based tool that runs in any modern browser. It connects to a Web server that stores our indices and handles search requests (via our Solr search server). We have developed a series of scripts to retrieve MEDLINE updates from the NLM public ftp site and to construct the indices for MEDLINE and HPO in our Lucene index. We have also developed an application programming interface (API) that handles requests for searches and other basic functions. The front-end has been developed using HTML5, CSS, and JavaScript. The visualizations have been developed using D3.js [54], a popular JavaScript visualization library. Figure 2 provides a diagrammatic overview of the architecture of the OVERT-MED system.

Figure 2. Client-server architecture of the Ontology-Driven Visual Search and Triage Interface for MEDLINE (OVERT-MED) system.

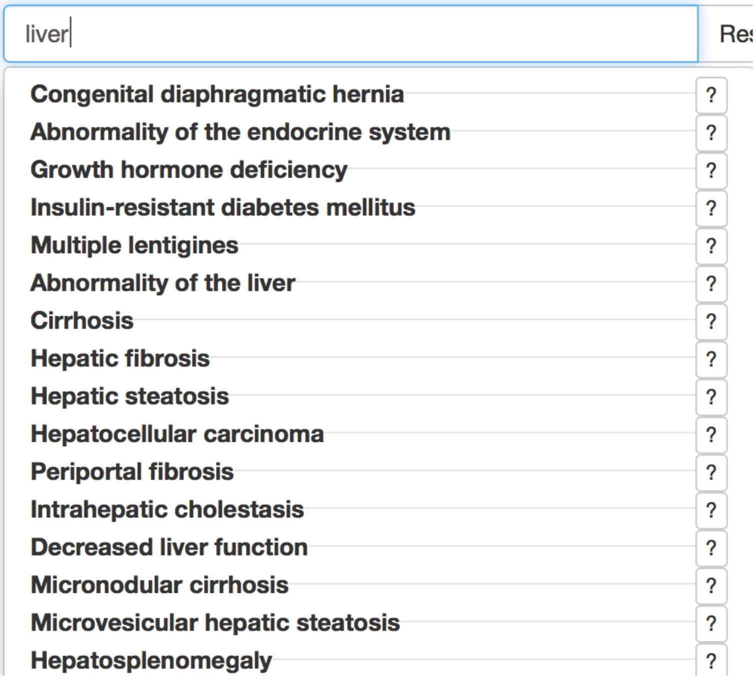


Results

Ontology Term Suggestion

OVERT-MED uses HPO to help users better articulate their search needs through a technique we call *ontology term suggester*. Users enter terms into a text box, and a set of suggestions (phenotypes) are provided. The suggestions are updated in real-time as a user types each character. In addition, to providing better terminological support, we look for matches on both the phenotype names as well as descriptions and expert commentary on the phenotypes (these are not shown to users, but are indexed on our server). For example, a user may be interested in finding articles related to the term “liver,” but may not have sufficient vocabulary to articulate a useful query involving relevant terms. Figure 3 shows the ontology term suggester after typing “liver” into the search box. Phenotypes related to the liver are displayed. Results such as “Growth hormone deficiency” and “Ascites” are displayed because they have a connection to the liver—the effects of growth hormone are mediated by insulin-like growth factor, which is produced primarily in the liver; and ascites is commonly associated with liver disease. Many of the returned phenotypes do not have the term *liver* in their name, but are related to the liver. In a traditional search interface, there is no way for a user to get from “liver” to “ascites” or “growth hormone deficiency.” Finally, because users may not understand a particular phenotype (eg, congenital diaphragmatic hernia), selecting the “?” button will open a new tab and load the official entry in the HPO Web browser. From there users can find more details, including associated genes and diseases. This search strategy can help users build knowledge of the domain and vocabulary that can be used to enhance cognitive performance and exploration.

Figure 3. The ontology term suggester, showing results from typing “liver.”



Congenital diaphragmatic hernia ?
Abnormality of the endocrine system ?
Growth hormone deficiency ?
Insulin-resistant diabetes mellitus ?
Multiple lentigines ?
Abnormality of the liver ?
Cirrhosis ?
Hepatic fibrosis ?
Hepatic steatosis ?
Hepatocellular carcinoma ?
Periportal fibrosis ?
Intrahepatic cholestasis ?
Decreased liver function ?
Micronodular cirrhosis ?
Microvesicular hepatic steatosis ?
Hepatosplenomegaly ?

Sensitivity Encoding for Query Refinement

A well-known problem in open-ended search tasks is that potentially relevant results may not be displayed if they do not meet the specified search criteria. For example, when searching for a house to buy, users often have ill-formed criteria, such as price range, number of bedrooms and bathrooms, yard size, location, and so on. Although certain search criteria may be specified (eg, 4 bedrooms, under $200,00), results that do not meet the criteria may also be relevant, such as a house that has only 3 bedrooms but is a great price. When using visualizations to support such search tasks, certain criteria can be relaxed and results that do not meet certain criteria can be visually encoded in different ways. For instance, results that do not meet number of bedrooms can be encoded with 1 color; results that do not meet yard size can be encoded with another; and so on. Visually encoding this type of information can provide cues to users to adjust their search criteria so that potentially relevant results are included. This visualization strategy, known as sensitivity encoding, has been shown to be beneficial in a number of contexts [55,56].

Although OVERT-MED supports the selection of precise phenotype names, the exact combination of words in a name may be too restrictive, and may not provide the most relevant results. For example, a user may select the phenotype *progressive external ophthalmoplegia*. Our index shows 811 articles associated with this specific phenotype. However, users may be interested in articles associated with different variations of the words—for example, *progressive opthalmoplegia* or *external opthalmoplegia.* We use a set of *Sensitivity Encoded Query Selectors* in OVERT-MED to handle this issue. When a phenotype is selected, we perform searches on our index using all possible combinations of the words and then visually encode the size of the result set. Figure 4 shows the result of a user selecting "progressive external opthalmoplegia." The number of matching articles for each combination is provided numerically and encoded visually using the length of the bar next to each combination. From Figure 4, we can see that if the user relaxes the term to "progressive ophthalmoplegia," an additional 104 articles show up in the index and with "external opthalmoplegia," an additional 418 articles show up. Without such a sensitivity encoding strategy, many of these potentially relevant results would not be made available. As users are often interested in more than 1 phenotype, multiple phenotypes can be selected, each of which is subjected to the same sensitivity encoding process. Figure 5 shows a second phenotype, congenital fibrosis of extraocular muscles, being added.

Figure 4. A set of sensitivity-encoded query selectors for "progressive external ophthalmoplegia."

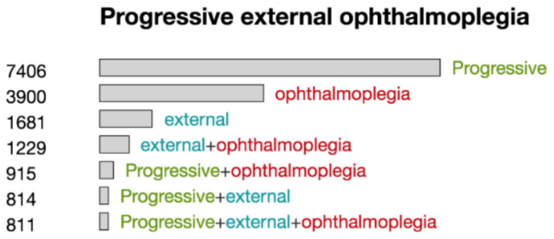
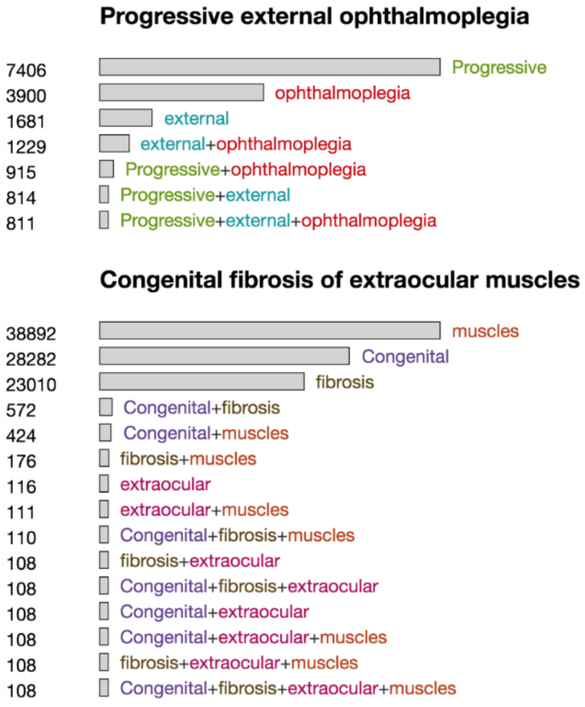


Figure 5. The result of adding a second phenotype via the ontology term suggester, which leads to more sensitivity-encoded query selectors.



Interactive Triaging Support to Mitigate Information Overload

OVERT-MED provides multistage triaging support to mitigate the information overload problem. Multiple design strategies support the first stage of triaging—the “multiple document” stage. First, when a specific set of terms is chosen, the metadata from up to 250 documents are visualized. Each document is encoded using a small bar, and the presence of each term is encoded using a section of the bar. Figure 6 shows how 6 documents are represented in the case of 3 terms (progressive external opthalmoplegia). Within the visualization, each row represents a document, and each column represents one of the phenotype words. The words are color coded—in this case, green for progressive, teal for external, and red for opthalmoplegia. A white cell indicates no occurrence of the word. The visualization functions as a type of heatmap [57], where the color saturation encodes the frequency of a term within a document. We call this technique the *query result heatmap*. In Figure 6, a darker red means higher occurrence of the word opthalmoplegia. This type of encoding can aid in rapid visual scanning and identification of potentially relevant documents [43,58].

To further support the triaging activity, OVERT-MED allows users to interactively explore metadata associated with the matching documents. Figure 7 shows the state of the interface after a user has selected “progressive+opthalmoplegia.” The first 250 documents (ranked by our indexing algorithm) are encoded in the Query Result Heatmap. Each row functions as an individual document heatmap, showing the occurrence of the 7 phenotype terms within the document. Because the user has selected “progressive” and “opthalmoplegia,” all documents indicate occurrences of both terms. It is readily apparent that most of the documents also contain the term “external.” Approximately 20 also contain “muscles,” 4 contain “extraocular,” 1 contains “fibrosis,” and 1 “congenital.”

OVERT-MED also provides a *Term Distribution Matrix* to help users quickly determine document relevance while browsing the Query Result Heatmap. Within the term distribution matrix, users can see the occurrence of terms in 4 places within the document metadata: (1) title, (2) journal name, (3) MeSH terms, and (4) abstract text. The document title, journal, year, and MeSH terms are also displayed. This representation helps users make decisions about relevance via quick visual scanning. For example, if a term appears only in the journal name it may not be very relevant, but if a term appears 5 times in the abstract text it is more likely to be relevant. Users can perceive this type of information quickly due to the categorical color encodings. Figure 8 shows the term distribution matrix for 2 different documents within the same result set. Through rapid visual scanning, even without reading the text, it is apparent that the terms are quite important in the document on the right.

To support rapid exploration—a fundamental goal of triaging—the keyboard arrow keys can be used to move quickly through the documents while the metadata is dynamically updated. If a relevant document is detected, users can hit the “enter” key or click the button to add the document to a pile for subsequent investigation (this stage is explained in greater detail in the following). This stage of triaging also allows for quick comparison of cooccurring phenotypes within documents. For example, Figure 9 shows the result of a user adding documents containing “congenital” and “fibrosis.” It is immediately clear through quick visual scanning that not many documents contain both “congenital fibrosis” and “opthalmoplegia.”

While browsing the query result heatmap, it may be difficult to remember which documents have been visited previously. This is especially true in the context of iterative triaging, where users may return to the heatmap after being away for some time. In OVERT-MED, when users pause on a document for 5 s or more, a small mark is placed beside the document to serve as a visual reminder (Figure 10). When revisiting the heatmap, users can quickly recognize which documents they have previously examined. We assume that 5 s is a reasonable threshold for determining when a user has examined the *term distribution matrix*.

Figure 6. The query result heatmap: 6 documents are represented by 6 rows, where each column represents a term (progressive external opthalmoplegia).

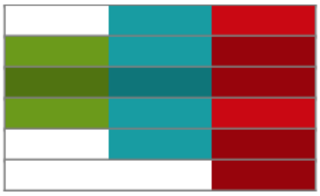
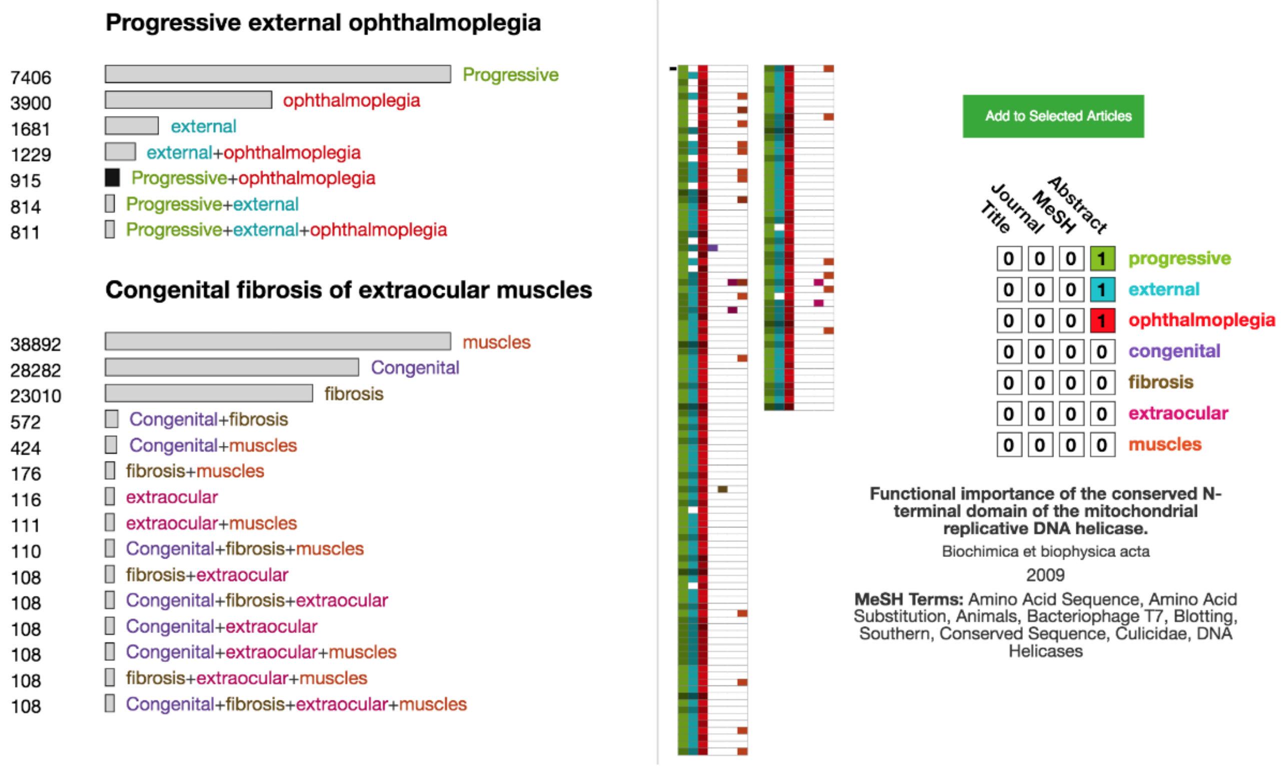


Figure 7. State of the interface after a user has selected “progressive+opthalmoplegia.”



Progressive external ophthalmoplegia

7406		Progressive
3900		ophthalmoplegia
1681		external
1229		external+ophthalmoplegia
915	■	Progressive+ophthalmoplegia
814	□	Progressive+external
811	□	Progressive+external+ophthalmoplegia

Congenital fibrosis of extraocular muscles

38892		muscles
28282		Congenital
23010		fibrosis
572	□	Congenital+fibrosis
424	□	Congenital+muscles
176	□	fibrosis+muscles
116	□	extraocular
111	□	extraocular+muscles
110	□	Congenital+fibrosis+muscles
108	□	fibrosis+extraocular
108	□	Congenital+fibrosis+extraocular
108	□	Congenital+extraocular
108	□	Congenital+extraocular+muscles
108	□	fibrosis+extraocular+muscles
108	□	Congenital+fibrosis+extraocular+muscles

Add to Selected Articles

Title	Journal	MeSH	Abstract	
0	0	0	1	progressive
0	0	0	1	external
0	0	0	1	ophthalmoplegia
0	0	0	0	congenital
0	0	0	0	fibrosis
0	0	0	0	extraocular
0	0	0	0	muscles

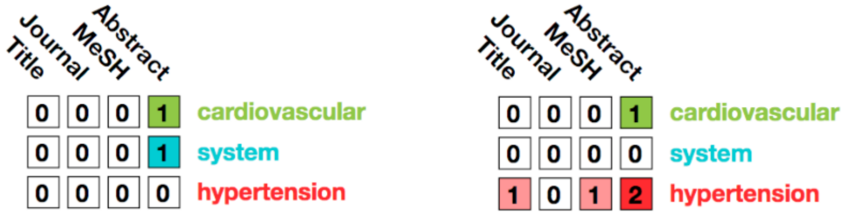
Functional importance of the conserved N-terminal domain of the mitochondrial replicative DNA helicase.

Biochimica et biophysica acta

2009

MeSH Terms: Amino Acid Sequence, Amino Acid Substitution, Animals, Bacteriophage T7, Blotting, Southern, Conserved Sequence, Culicidae, DNA Helicases

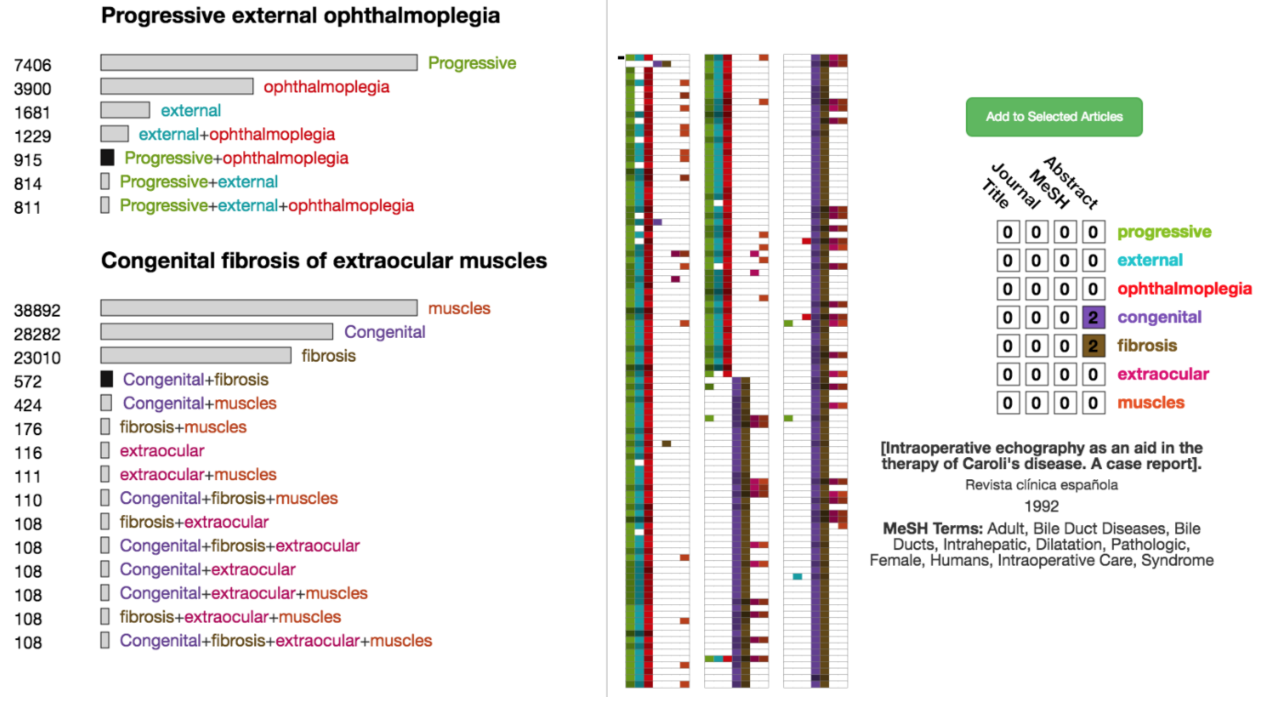
Figure 8. The term distribution matrix for 2 different documents within the same result set.



Title	Journal	MeSH	Abstract	
0	0	0	1	cardiovascular
0	0	0	1	system
0	0	0	0	hypertension

Title	Journal	MeSH	Abstract	
0	0	0	1	cardiovascular
0	0	0	0	system
1	0	1	2	hypertension

Figure 9. The result of a user adding documents containing “congenital” and “fibrosis” for comparison.



Progressive external ophthalmoplegia

7406		Progressive
3900		ophthalmoplegia
1681		external
1229		external+ophthalmoplegia
915	■	Progressive+ophthalmoplegia
814	□	Progressive+external
811	□	Progressive+external+ophthalmoplegia

Congenital fibrosis of extraocular muscles

38892		muscles
28282		Congenital
23010		fibrosis
572	■	Congenital+fibrosis
424	□	Congenital+muscles
176	□	fibrosis+muscles
116	□	extraocular
111	□	extraocular+muscles
110	□	Congenital+fibrosis+muscles
108	□	fibrosis+extraocular
108	□	Congenital+fibrosis+extraocular
108	□	Congenital+extraocular
108	□	Congenital+extraocular+muscles
108	□	fibrosis+extraocular+muscles
108	□	Congenital+fibrosis+extraocular+muscles

Add to Selected Articles

Title	Journal	MeSH	Abstract	
0	0	0	0	progressive
0	0	0	0	external
0	0	0	0	ophthalmoplegia
0	0	0	2	congenital
0	0	0	2	fibrosis
0	0	0	0	extraocular
0	0	0	0	muscles

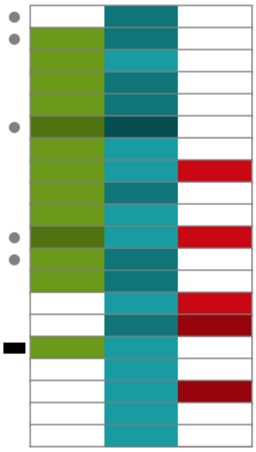
[Intraoperative echography as an aid in the therapy of Caroli's disease. A case report].

Revista clínica española

1992

MeSH Terms: Adult, Bile Duct Diseases, Bile Ducts, Intrahepatic, Dilatation, Pathologic, Female, Humans, Intraoperative Care, Syndrome

Figure 10. Closeup view of the query result heatmap.

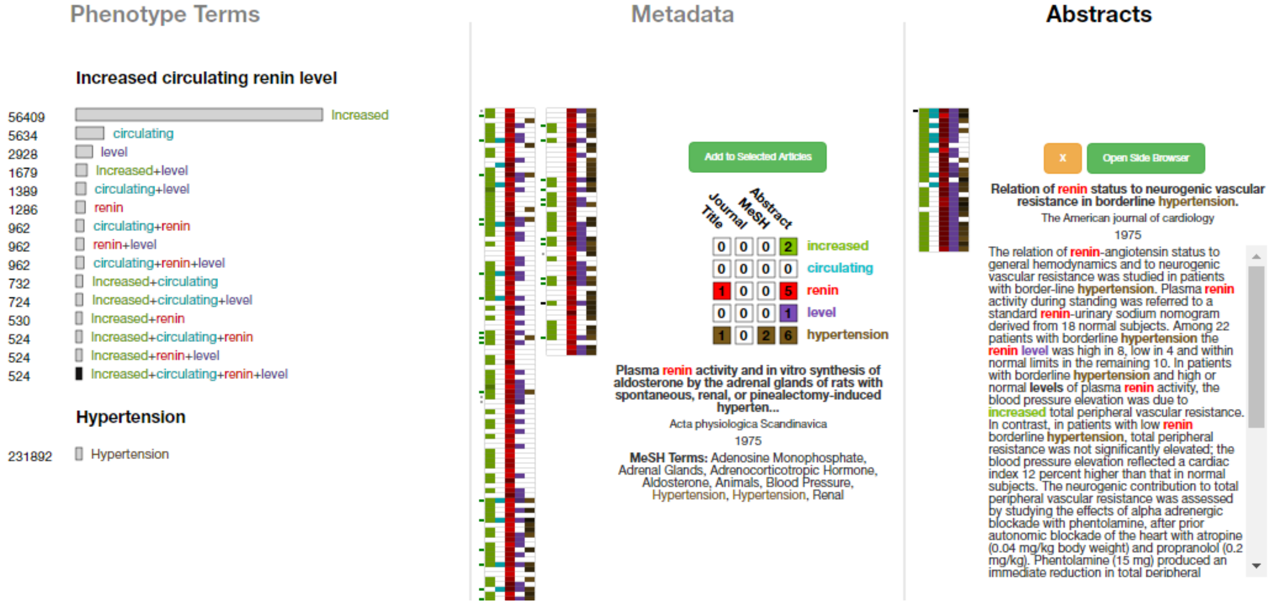


The next stage in the triaging activity—the "individual document" stage—involves examining individual abstracts of previously chosen articles. At this stage, users are likely to have narrowed down the number of documents significantly. Documents are encoded via a *Selected Pile Heatmap* in the same manner as in the query result heatmap, and each can be selected to view its abstract. In this *term-encoded abstract*, matching terms are color coded to facilitate quick identification, especially within the abstract text. Figure 11 shows an example in which the user has selected 29 documents, which are encoded in the selected pile heatmap and the term-encoded abstract is displayed for the first document. Even before reading the text in detail, it is easy to see that "renin" and "hypertension" both appear frequently, indicating that they are important. Thus, users can scan the text quickly to get a sense of the appearance of the query terms, without having to necessarily read the text sequentially. An important aspect of this stage of triaging is the ability to quickly categorize documents. In OVERT-MED, users can quickly reject a paper by selecting the orange "x" button, or can quickly add a paper to the next stage by selecting the green button or pressing the "enter" key.

The final stage of triaging is the "further reading" stage, where a small set of documents are read in-depth to extract relevant information and satisfy the original information need. Although this stage could be supported in various ways, we support this stage in OVERT-MED by presenting a PubMed entry for a selected document in an embedded frame directly within the interface of OVERT-MED. This allows for quick inspection of any PubMed details that are important to the user, such as full-text links, citation details, and PubMed Commons links, and also allows users to login to their *National Center for Biotechnology Information* (NCBI) account to save the article to a collection, compare with other saved articles, and so on. There is also a button to open the PubMed link in a new browser tab if a user needs more space. Figure 12 shows a full-screen capture of OVERT-MED in which a user has traversed all stages of a search and triaging activity.

As research shows that triaging activities are cyclical and iterative, we have designed OVERT-MED to be flexible in this regard. At any point during an activity, users may adjust their query or document selections, and each component of the interface will dynamically reflect any changes. For example, a user may reach the final stage of triaging and find a term within a document that seems relevant to the original information need. The user can return to the initial stage of entering the term and selecting phenotypes. In doing so, the rest of the interface remains stable and the user can proceed through any of the triaging stages. Figure 13 shows the interface after a user has examined a document in detail in the final stage, discovered a link between renin level (the original phenotype of interest) and arterial pressure, and has returned to the initial stage to find a phenotype related to arterial pressure. The user discovers a phenotype named "elevated mean arterial pressure" and selects it. At this stage, the user is not particularly interested in whether the arterial pressure is elevated, and simply wants to explore the relationship between renin level and arterial pressure. Due to our sensitivity encoding strategy, the user can select "arterial+pressure" to add documents with those 2 terms. From this point, the user can continue through the triaging stages or return to the initial stage again.

Figure 11. Twenty-nine documents have been selected to examine in closer detail.



Phenotype Terms

Increased circulating renin level

56409	□	Increased
5634	□	circulating
2928	□	level
1679	□	Increased+level
1389	□	circulating+level
1286	□	renin
962	□	circulating+renin
962	□	renin+level
962	□	circulating+renin+level
732	□	Increased+circulating
724	□	Increased+circulating+level
530	□	Increased+renin
524	□	Increased+circulating+renin
524	□	Increased+renin+level
524	■	Increased+circulating+renin+level

Hypertension

231892	□	Hypertension

Metadata

Title	Journal	MeSH	Abstract	
0	0	0	2	increased
0	0	0	0	circulating
1	0	0	5	renin
0	0	0	1	level
1	0	2	6	hypertension

Plasma renin activity and in vitro synthesis of aldosterone by the adrenal glands of rats with spontaneous, renal, or pinealectomy-induced hyperten...

Acta physiologica Scandinavica

1975

MeSH Terms: Adenosine Monophosphate, Adrenal Glands, Adrenocorticotropic Hormone, Aldosterone, Animals, Blood Pressure, Hypertension, Hypertension, Renal

Abstracts

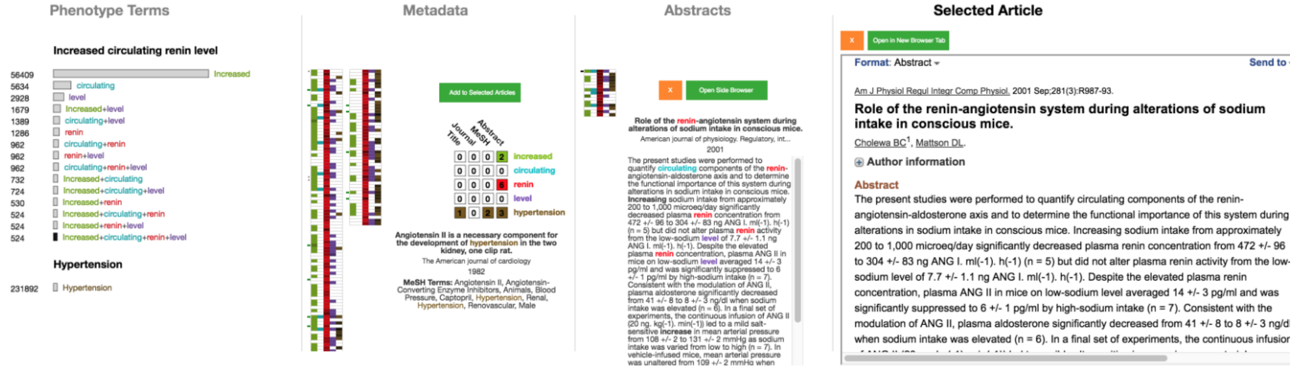
Relation of renin status to neurogenic vascular resistance in borderline hypertension.

The American journal of cardiology

1975

The relation of renin-angiotensin status to general hemodynamics and to neurogenic vascular resistance was studied in patients with border-line hypertension. Plasma renin activity during standing was referred to a standard renin-urinary sodium nomogram derived from 18 normal subjects. Among 22 patients with borderline hypertension the renin level was high in 8, low in 4 and within normal limits in the remaining 10. In patients with borderline hypertension and high or normal levels of plasma renin activity, the blood pressure elevation was due to increased total peripheral vascular resistance. In contrast, in patients with low renin borderline hypertension, total peripheral resistance was not significantly elevated; the blood pressure elevation reflected a cardiac index 12 percent higher than that in normal subjects. The neurogenic contribution to total peripheral vascular resistance was assessed by studying the effects of alpha adrenergic blockade with phentolamine, after prior autonomic blockade of the heart with atropine (0.04 mg/kg body weight) and propranolol (0.2 mg/kg). Phentolamine (15 mg) produced an immediate reduction in total peripheral

Figure 12. Full-screen capture showing all components of OVERT-MED where a user has traversed all stages of a search and triaging activity. OVERT-MED: Ontology-Driven Visual Search and Triage Interface for MEDLINE.



Phenotype Terms

Increased circulating renin level

56409	□	Increased
5634	□	circulating
2928	□	level
1679	□	Increased+level
1389	□	circulating+level
1286	□	renin
962	□	circulating+renin
962	□	renin+level
962	□	circulating+renin+level
732	□	Increased+circulating
724	□	Increased+circulating+level
530	□	Increased+renin
524	□	Increased+circulating+renin
524	□	Increased+renin+level
524	■	Increased+circulating+renin+level

Hypertension

231892	□	Hypertension

Metadata

Title	Journal	MeSH	Abstract	
0	0	0	2	increased
0	0	0	0	circulating
0	0	0	6	renin
0	0	0	0	level
1	0	2	3	hypertension

Angiotensin II is a necessary component for the development of hypertension in the two kidney, one clip rat.

The American journal of cardiology

1982

MeSH Terms: Angiotensin II, Angiotensin-Converting Enzyme Inhibitors, Animals, Blood Pressure, Captopril, Hypertension, Renal, Hypertension, Renovascular, Male

Abstracts

Role of the renin-angiotensin system during alterations of sodium intake in conscious mice.

American journal of physiology. Regulatory, int...

2001

The present studies were performed to quantify circulating components of the renin-angiotensin-aldosterone axis and to determine the functional importance of this system during alterations in sodium intake in conscious mice. Increasing sodium intake from approximately 200 to 1,000 microeq/day significantly decreased plasma renin concentration from 472 +/- 96 to 304 +/- 83 ng ANG I. ml(-1). h(-1) (n = 5) but did not alter plasma renin activity from the low-sodium level of 7.7 +/- 1.1 ng ANG I. ml(-1). h(-1). Despite the elevated plasma renin concentration, plasma ANG II in mice on low-sodium level averaged 14 +/- 3 pg/ml and was significantly suppressed to 6 +/- 1 pg/ml by high-sodium intake (n = 7). Consistent with the modulation of ANG II, plasma aldosterone significantly decreased from 41 +/- 8 to 8 +/- 3 ng/dl when sodium intake was elevated (n = 6). In a final set of experiments, the continuous infusion of ANG II (20 ng. kg(-1). min(-1)) led to a mild salt-sensitive increase in mean arterial pressure from 108 +/- 2 to 131 +/- 2 mmHg as sodium intake was varied from low to high (n = 7). In vehicle-infused mice, mean arterial pressure was unaltered from 109 +/- 2 mmHg when

Selected Article

Am J Physiol Regul Integr Comp Physiol. 2001 Sep;281(3):R987-93.

Role of the renin-angiotensin system during alterations of sodium intake in conscious mice.

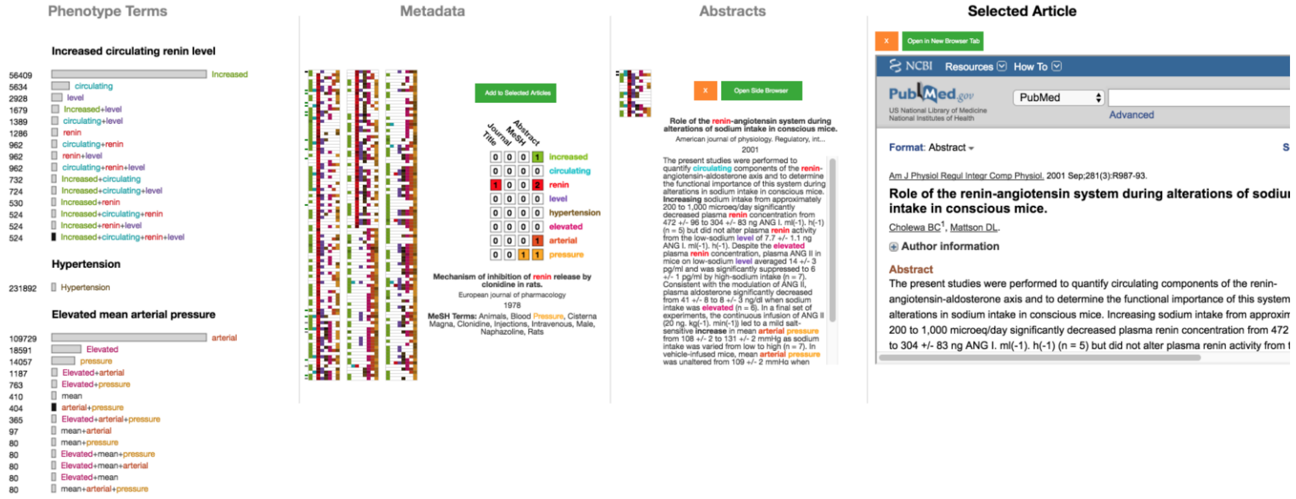
Cholewa BC[1], Mattson DL.

⊕ Author information

Abstract

The present studies were performed to quantify circulating components of the renin-angiotensin-aldosterone axis and to determine the functional importance of this system during alterations in sodium intake in conscious mice. Increasing sodium intake from approximately 200 to 1,000 microeq/day significantly decreased plasma renin concentration from 472 +/- 96 to 304 +/- 83 ng ANG I. ml(-1). h(-1) (n = 5) but did not alter plasma renin activity from the low-sodium level of 7.7 +/- 1.1 ng ANG I. ml(-1). h(-1). Despite the elevated plasma renin concentration, plasma ANG II in mice on low-sodium level averaged 14 +/- 3 pg/ml and was significantly suppressed to 6 +/- 1 pg/ml by high-sodium intake (n = 7). Consistent with the modulation of ANG II, plasma aldosterone significantly decreased from 41 +/- 8 to 8 +/- 3 ng/dl when sodium intake was elevated (n = 6). In a final set of experiments, the continuous infusior

Figure 13. The interface after a user has examined a document in detail in the final stage, discovered a link, and has returned to the initial stage with a new information need.



Phenotype Terms

Increased circulating renin level

56409	□	Increased
5634	□	circulating
2928	□	level
1679	□	Increased+level
1389	□	circulating+level
1286	□	renin
962	□	circulating+renin
962	□	renin+level
962	□	circulating+renin+level
732	□	Increased+circulating
724	□	Increased+circulating+level
530	□	Increased+renin
524	□	Increased+circulating+renin
524	□	Increased+renin+level
524	■	Increased+circulating+renin+level

Hypertension

231892	□	Hypertension

Elevated mean arterial pressure

109729	□	arterial
18591	□	Elevated
14057	□	pressure
1187	□	Elevated+arterial
763	□	Elevated+pressure
410	□	mean
404	■	arterial+pressure
365	□	Elevated+arterial+pressure
97	□	mean+arterial
80	□	mean+pressure
80	□	Elevated+mean+pressure
80	□	Elevated+mean+arterial
80	□	Elevated+mean
80	□	mean+arterial+pressure

Metadata

Title	Journal	MeSH	Abstract	
0	0	0	1	increased
0	0	0	0	circulating
1	0	0	2	renin
0	0	0	0	level
0	0	0	0	hypertension
0	0	0	0	elevated
0	0	0	1	arterial
0	0	1	1	pressure

Mechanism of inhibition of renin release by clonidine in rats.

European journal of pharmacology

1978

MeSH Terms: Animals, Blood Pressure, Cisterna Magna, Clonidine, Injections, Intravenous, Male, Naphazoline, Rats

Abstracts

Role of the renin-angiotensin system during alterations of sodium intake in conscious mice.

American journal of physiology. Regulatory, int...

2001

The present studies were performed to quantify circulating components of the renin-angiotensin-aldosterone axis and to determine the functional importance of this system during alterations in sodium intake in conscious mice. Increasing sodium intake from approximately 200 to 1,000 microeq/day significantly decreased plasma renin concentration from 472 +/- 96 to 304 +/- 83 ng ANG I. ml(-1). h(-1) (n = 5) but did not alter plasma renin activity from the low-sodium level of 7.7 +/- 1.1 ng ANG I. ml(-1). h(-1). Despite the elevated plasma renin concentration, plasma ANG II in mice on low-sodium level averaged 14 +/- 3 pg/ml and was significantly suppressed to 6 +/- 1 pg/ml by high-sodium intake (n = 7). Consistent with the modulation of ANG II, plasma aldosterone significantly decreased from 41 +/- 8 to 8 +/- 3 ng/dl when sodium intake was elevated (n = 6). In a final set of experiments, the continuous infusion of ANG II (20 ng. kg(-1). min(-1)) led to a mild salt-sensitive increase in mean arterial pressure from 108 +/- 2 to 131 +/- 2 mmHg as sodium intake was varied from low to high (n = 7). In vehicle-infused mice, mean arterial pressure was unaltered from 109 +/- 2 mmHg when

Selected Article

Am J Physiol Regul Integr Comp Physiol. 2001 Sep;281(3):R987-93.

Role of the renin-angiotensin system during alterations of sodiu intake in conscious mice.

Cholewa BC[1], Mattson DL.

⊕ Author information

Abstract

The present studies were performed to quantify circulating components of the renin-angiotensin-aldosterone axis and to determine the functional importance of this system alterations in sodium intake in conscious mice. Increasing sodium intake from approxim 200 to 1,000 microeq/day significantly decreased plasma renin concentration from 472 to 304 +/- 83 ng ANG I. ml(-1). h(-1) (n = 5) but did not alter plasma renin activity from t

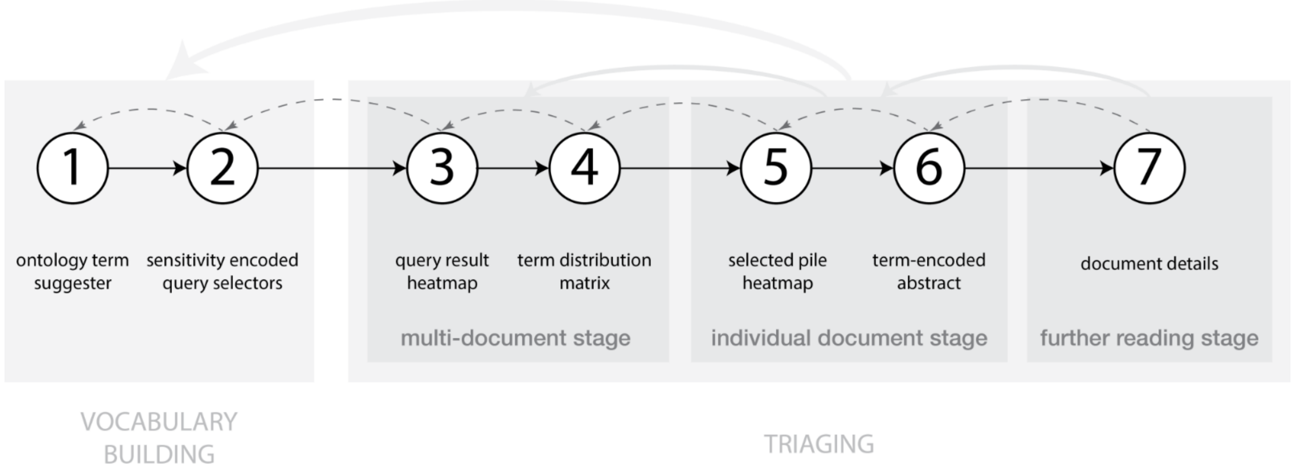
Discussion

Overview

OVERT-MED was developed to address two major problems that are known to exist in complex, exploratory search activities: (1) the difficulty in articulating information needs due to insufficient knowledge and domain-specific vocabulary, and (2) the difficultly in dealing with information overload due to the large number of results returned. To address the first difficulty, we proposed the idea of using a formal ontology to help users build domain-specific terminology and knowledge for constructing search queries. To assist in this process, we indexed HPO and provided a search feature that provides robust results to terms that are entered. To address the problem of search criteria being too restrictive in open-ended contexts, we used a visual sensitivity encoding strategy to help users see possibilities with different combinations of terms.

There are 7 main steps that users take when performing search and triaging tasks with OVERT-MED—the first 2 within a vocabulary building phase and the next 5 within a triaging phase. The triaging phase can be broken down into the 3 key stages. Figure 14 provides an overview of this process and shows the techniques we use to help users at each step. To help users build vocabulary and generate queries, we use an *ontology term suggester* and *sensitivity encoded query selectors*. After selecting a query, users move to the triaging phase, where they traverse through 3 stages. During the first stage—the multi-document stage—users are presented with a *query result heatmap* that encodes the appearance and frequency of query terms within the document result set. A keyboard interaction technique enables rapid navigation through the documents. To facilitate assessment at this stage, a *term distribution matrix* provides more information about each document within the heatmap. Together these techniques allow for rapid scanning to assess relevance and select documents for the next stage. During the second triaging stage—the individual document stage—users are presented with a *Selected Pile Heatmap* that encodes only the selected documents from the previous stage. As users browse the heatmap, they can inspect a *term-encoded abstract* of each individual document. The term-encoding supports quick detection of the appearance of query terms within the document abstract. After assessing the relevance of individual documents, users select documents to move to the next stage. During the third triaging stage—the further reading stage—users focus on a single document by viewing details in depth. Here, the PubMed entry for a document can be retrieved directly within OVERT-MED or within a new browser tab. At any point in the overall activity, users can return to any step and continue from there, which supports the iterative and cyclical nature of search and triaging tasks.

Figure 14. Overall search and triage process supported by OVERT-MED. Users take 7 main steps—the first 2 within a vocabulary building phase, and the next 5 within a triaging phase. OVERT-MED: Ontology-Driven Visual Search and Triage Interface for MEDLINE.



Validation

Ongoing formative evaluation suggests that the design features in OVERT-MED can mitigate the two problems mentioned above. We tested OVERT-MED with a small group of users who are not domain-experts, and our proposal to use a formal ontology to help users articulate their information needs does seem to be useful. As mentioned previously, different types of users are known to search the scientific literature, many of which are not domain experts. For example, pediatricians often try to identify abnormal phenotypes in patients before referring them to a clinical geneticist. However, because they are not domain experts, pediatricians may not have very extensive knowledge and vocabulary of phenotypes. Even if they search the literature to identify phenotype names (eg, via PubMed), they may still not find phenotypes that are related to one another. As another example, patients are known to search the literature to learn more about their own conditions. As they are not domain experts, patients could also benefit from having access to an ontology such as HPO to help them build domain-specific knowledge and vocabulary. Thus, testing with users who are not domain experts can give an indication of the usefulness of our design strategies.

In our testing, we noticed that although an ontology can help users develop more appropriate vocabulary, users do not necessarily develop a good understanding of the ontology itself. As a robust mental model of the ontology may lead to even better search performance (eg, by knowing which entities are highly connected to others, knowing relationships among entities at multiple levels of abstraction, and so on), we have decided to pursue a solution to this as future work (see Future Work section). In addition, our multistage triaging shows promise in mitigating the information overload problem. Users were able to go back and forth through the triaging stages to satisfy information needs without being overwhelmed by long lists of documents.

Limitations

There is 1 current limitation of OVERT-MED that should be noted: the MEDLINE data are limited to metadata and abstract text only, and do not include full texts. This is simply because the NLM does not release full-texts due to copyright issues. There is little we can do to address this issue. Empirical evidence, however, does suggest that the document title and abstract are among the most important features of a document in determining its relevance [37], so perhaps it is not a critical limitation.

Future Work

We envision at least three lines of valuable future research:

First, developing interactive visualization techniques to support ontology sensemaking. The intention behind the current version of OVERT-MED is to help address the common problem of lack of adequate vocabulary. Although OVERT-MED appears to support users in improving their search terms and potentially developing some domain knowledge, it does not necessarily support users in making sense of the ontology itself—that is, understanding its size, organization, types of relationships, significant and insignificant entities, and so on. Interactive visualizations of ontologies may enhance search and triaging activities. Second, testing OVERT-MED with different ontologies in different contexts. This will help assess the transferability of the design features of OVERT-MED. Third, conducting formal testing of OVERT-MED. Although our informal testing has been useful, more formal testing will provide validation of the design strategies.

Conclusions

We have developed a Web-based interactive visualization tool, OVERT-MED, to address two common problems in exploratory search—namely, the lack of adequate vocabulary to construct useful queries and the difficulty of dealing with very large result sets. The novelty of our approach is in the combination of (1) using an ontology to help build domain-specific knowledge and vocabulary when users need to be kept “in the loop” and (2) providing multistage triaging support using interactive visualizations to help mitigate the information overload problem. We anticipate these ideas can be applied successfully in other contexts where either of these issues exists.

Acknowledgments

This research has been supported financially by the Natural Sciences and Engineering Council of Canada.

Conflicts of Interest

None declared.

References

1. Krupski TL, Dahm P, Fesperman SF, Schardt CM. How to perform a literature search. J Urol 2008 Apr;179(4):1264-1270. [doi: 10.1016/j.juro.2007.11.087] [Medline: 18280516]
2. Islamaj DR, Murray GC, Névéol A, Lu Z. Understanding PubMed user search behavior through log analysis. Database (Oxford) 2009 Nov 27;2009:bap018 [FREE Full text] [doi: 10.1093/database/bap018] [Medline: 20157491]
3. Kritz M, Gschwandtner M, Stefanov V, Hanbury A, Samwald M. Utilization and perceived problems of online medical resources and search tools among different groups of European physicians. J Med Internet Res 2013 Jun 26;15(6):e122 [FREE Full text] [doi: 10.2196/jmir.2436] [Medline: 23803299]
4. Hersh WR, Crabtree MK, Hickam DH, Sacherek L, Friedman CP, Tidmarsh P, et al. Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions. J Am Med Inform Assoc 2002;9(3):283-293 [FREE Full text] [Medline: 11971889]
5. Parsons P, Sedig K, Mercer R, Khordad M, Knoll J, Rogan P. Visual analytics for supporting evidence-based interpretation of molecular cytogenomic findings. In: Proceedings of the 2015 Workshop on Visual Analytics in Healthcare. New York, New York, USA: ACM Press; 2015 Oct 25 Presented at: IEEE VIS; 2015; Chicago. [doi: 10.1145/2836034.2836036]
6. Palotti J, Hanbury A, Müller H, Kahn C. How users search and what they search for in the medical domain. Inf Retrieval J 2015 Oct 24;19(1-2):189-224. [doi: 10.1007/s10791-015-9269-8]
7. Marchionini G. Exploratory search: from finding to understanding. In: Communications of the ACM - Supporting exploratory search. New York, NY, USA: ACM; Apr 01, 2006:41-46.
8. Hersh WR, Hickam DH. How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review. J Am Med Assoc 1998 Oct 21;280(15):1347-1352. [Medline: 9794316]
9. Cui L, Carter R, Zhang G. Evaluation of a novel conjunctive exploratory navigation interface for consumer health information: a crowdsourced comparative study. J Med Internet Res 2014;16(2):e45 [FREE Full text] [doi: 10.2196/jmir.3111] [Medline: 24513593]
10. Pang PC, Chang S, Verspoor K, Pearce J. Designing health websites based on users' web-based information-seeking behaviors: a mixed-method observational study. J Med Internet Res 2016 Jun 06;18(6):e145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]
11. Ely JW, Osheroff JA, Chambliss ML, Ebell MH, Rosenbaum ME. Answering physicians' clinical questions: obstacles and potential solutions. J Am Med Inform Assoc 2005;12(2):217-224 [FREE Full text] [doi: 10.1197/jamia.M1608] [Medline: 15561792]

12. Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. Health Info Libr J 2007 Jun;24(2):78-94 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00713.x] [Medline: 17584211]
13. Dietze H, Alexopoulou D, Alvers MR, Barrio-Alvers L, Andreopoulos B, Doms A, et al. GoPubMed: Exploring PubMed with Ontological Background Knowledge. In: Bioinformatics for Systems Biology. Bioinforma Syst Biol Totowa, NJ: Humana Press; 2009:385-399.
14. NCBI.NLM. Home-PubMed URL: https://www.ncbi.nlm.nih.gov/pubmed [accessed 2016-10-14] [WebCite Cache ID 6lGAJxbQD]
15. Furnas GW, Landauer TK, Gomez LM, Dumais ST. The vocabulary problem in human-system communication. Commun ACM 1987:964-971.
16. Belkin NJ. Helping people find what they don't know. Commun ACM 2000;43(8):58-61. [doi: 10.1145/345124.345143]
17. Patrick TB, Monga HK, Sievert ME, Houston HJ, Longo DR. Evaluation of controlled vocabulary resources for development of a consumer entry vocabulary for diabetes. J Med Internet Res 2001;3(3):E24 [FREE Full text] [doi: 10.2196/jmir.3.3.e24] [Medline: 11720966]
18. Plovnick RM, Zeng QT. Reformulation of consumer health queries with professional terminology: a pilot study. J Med Internet Res 2004 Sep 03;6(3):e27 [FREE Full text] [doi: 10.2196/jmir.6.3.e27] [Medline: 15471753]
19. Sievert M, Patrick T, Reid J. Need a bloody nose be a nosebleed? or, lexical variants cause surprising results. Bull Med Libr Assoc 2001 Jan;89(1):68-71 [FREE Full text] [Medline: 11209803]
20. Zeng QT, Tse T. Exploring and developing consumer health vocabularies. J Am Med Inform Assoc 2006;13(1):24-29 [FREE Full text] [doi: 10.1197/jamia.M1761] [Medline: 16221948]
21. Lowe HJ, Barnett GO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. J Am Med Assoc 1994 Apr 13;271(14):1103-1108. [Medline: 8151853]
22. Malhotra A, Gündel M, Rajput AM, Mevissen H, Saiz A, Pastor X, et al. Knowledge retrieval from PubMed abstracts and electronic medical records with the Multiple Sclerosis Ontology. PLoS One 2015;10(2):e0116718 [FREE Full text] [doi: 10.1371/journal.pone.0116718] [Medline: 25665127]
23. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. Database (Oxford) 2011;2011:baq036 [FREE Full text] [doi: 10.1093/database/baq036] [Medline: 21245076]
24. Hoeber O, Khazaei T. Evaluating citation visualization and exploration methods for supporting academic search tasks. Online Information Review 2015 Apr 13;39(2):229-254. [doi: 10.1108/OIR-10-2014-0259]
25. Hoeber O. Visual Search Analytics: Combining Machine Learning and Interactive Visualization to Support Human-Centred Search. 2014 Presented at: Pro-ceedings of the MindTheGap'14 Workshop; March 4 2014; Berlin, Germany p. 37-43.
26. Hearst M, Elliott A, English J, Sinha R, Swearingen K, Yee K. Finding the flow in web site search. Commun ACM 2002;45(9):42-49. [doi: 10.1145/567498.567525]
27. Yee K, Swearingen K, Li K, Hearst M. Faceted metadata for image search and browsing. New York, NY, USA: ACM Press; 2003 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 5-10 2003; Ft Lauderdale, FL, USA p. 401-408. [doi: 10.1145/642611.642681]
28. Dork M, Williamson C, Carpendale S. Towards Visual Web Search?: Interactive Query Formulation and Search Result Visualization. 2009 Presented at: WSSP 2009: WWW Workshop on Web Search Result Summarization and Presentation; April 20 2009; Madrid, Spain p. 5.
29. Diriye A, Tombros A, Blandford A. A Little Interaction Can Go a Long Way: Enriching the Query Formulation Process. In: Lect Notes Comput Sci. 2012 Presented at: European Conference on Information Retrieval; April 1-5 2012; Barcelona, Spain p. 531-534. [doi: 10.1007/978-3-642-28997-2_57]
30. Joho H, Coverson C, Sanderson M, Beaulieu M. Hierarchical presentation of expansion terms. New York, NY, USA: ACM; 2002 Presented at: ACM symposium on Applied computing; March 11-14 2002; Madrid, Spain p. 645-649. [doi: 10.1145/508791.508916]
31. Gruber T. The role of common ontology in achieving sharable, reusable knowledge bases. 1991 Presented at: Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning; 1991; Cambridge, MA, USA.
32. Chandrasekaran B, Josephson JR, Benjamins VR. What are ontologies, and why do we need them? IEEE Intell Syst 1975 Oct 01;14(1):20-26.
33. Guarino N, Oberle D, Staab S. What Is an Ontology? In: Handbook on Ontologies. Handb Ontol Berlin, Heidelberg: Springer Berlin Heidelberg; 2009:1-17.
34. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet 2008 Nov;83(5):610-615 [FREE Full text] [doi: 10.1016/j.ajhg.2008.09.017] [Medline: 18950739]
35. Mavri A, Loizides F, Photiades T, Zaphiris P. We Have the Content…Now What?: The role of Structure and Interactivity in Academic Document Triage Interfaces. Inf Des J 2013;20(3):247-265. [doi: 10.1075/idj.20.3.05mav]
36. Loizides F, Buchanan G. Towards a Framework for Human (Manual) Information Retrieval. In: Multidisciplinary Information Retrieval. 2013 Presented at: Information Retrieval Facility Conference; October 7-9; Limassol, Cyprus p. 87-98. [doi: 10.1007/978-3-642-41057-4_10]

37. Loizides F, Buchanan G. An empirical study of user navigation during document triage. 2009 Presented at: 13th European Conference; September 27 - October 2; Corfu, Greece p. 138-149.

38. Khazaei T, Hoeber O. Supporting academic search tasks through citation visualization and exploration. Int J Digit Libr 2016 Apr 26:1-14. [doi: 10.1007/s00799-016-0170-x]

39. Spink A, Wolfram D, Jansen M, Saracevic T. Searching the web: the public and their queries. J Am Soc Inf Sci 2001;52(3):226-234. [doi: 10.1002/1097-4571(2000)9999:9999<::AID-ASI1591>3.0.CO;2-R]

40. Silverstein C, Marais H, Henzinger M, Moricz M. Analysis of a very large web search engine query log. SIGIR Forum 1999 Sep 01;33(1):6-12. [doi: 10.1145/331403.331405]

41. Scaife M, Rogers Y. External cognition: how do graphical representations work? Int J Hum Comput Stud 1996 Aug;45(2):185-213. [doi: 10.1006/ijhc.1996.0048]

42. Larkin J, Simon H. Why a diagram is (sometimes) worth ten thousand words. Cogn Sci 1987;11(1):65-100. [doi: 10.1111/j.1551-6708.1987.tb00863.x]

43. Hearst M. TileBars: Visualization of Term Distribution Information in Full Text Information Access. 1995 Presented at: Proc SIGCHI Conf Hum factors Comput Syst; 1995; Denver, CO, USA. [doi: 10.1145/223904.223912]

44. Gorg C, Liu Z, Stasko J. Reflections on the evolution of the Jigsaw visual analytics system. Inf Vis 2013 Jul 23;13(4):336-345. [doi: 10.1177/1473871613495674]

45. Thomas W, Alexopoulou D, Dietze H, Schroeder M. Searching biomedical literature with anatomy ontologies. Anatomy Ontologies for Bioinformatics 2009;6:177-194. [doi: 10.1007/978-1-84628-885-2_9]

46. Wei C, Kao H, Lu Z. PubTator: a web-based text mining tool for assisting biocuration. Nucleic Acids Res 2013 Jul;41(Web Server issue):W518-W522 [FREE Full text] [doi: 10.1093/nar/gkt441] [Medline: 23703206]

47. Muin M, Fontelo P, Liu F, Ackerman M. SLIM: an alternative Web interface for MEDLINE/PubMed searches - a preliminary study. BMC Med Inform Decis Mak 2005 Dec 01;5:37 [FREE Full text] [doi: 10.1186/1472-6947-5-37] [Medline: 16321145]

48. Eaton AD. HubMed: a web-based biomedical literature search interface. Nucleic Acids Res 2006 Jul 01;34(Web Server issue):W745-W747 [FREE Full text] [doi: 10.1093/nar/gkl037] [Medline: 16845111]

49. Wang J, Cetindil I, Ji S, Li C, Xie X, Li G, et al. Interactive and fuzzy search: a dynamic way to explore MEDLINE. Bioinformatics 2010 Sep 15;26(18):2321-2327 [FREE Full text] [doi: 10.1093/bioinformatics/btq414] [Medline: 20624778]

50. Yu H, Kim T, Oh J, Ko I, Kim S. RefMed: relevance feedback retrieval system fo PubMed. New York, NY, USA: ACM; 2009 Presented at: 18th ACM conference on Information and knowledge management; November 2-6 2009; Hong Kong, China p. 2099-2100. [doi: 10.1145/1645953.1646322]

51. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. Database (Oxford) 2011;2011:baq036 [FREE Full text] [doi: 10.1093/database/baq036] [Medline: 21245076]

52. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Inf Process Manag 1988 Jan;24(5):513-523. [doi: 10.1016/0306-4573(88)90021-0]

53. Salton G, Wong A, Yang CS. A vector space model for automatic indexing. Commun ACM 1975;18(11):613-620 [FREE Full text] [doi: 10.1145/361219.361220]

54. Bostock M, Ogievetsky V, Heer J. D^3: data-driven documents. IEEE Trans Vis Comput Graph 2011 Dec;17(12):2301-2309. [doi: 10.1109/TVCG.2011.185] [Medline: 22034350]

55. Spence R, Tweedie L. The Attribute Explorer: information synthesis via exploration. Interact Comput 1998 Dec;11(2):137-146. [doi: 10.1016/S0953-5438(98)00022-8]

56. Spence R. Sensitivity encoding to support information space navigation: a design guideline. Inf Vis 2002;1(2):120-129. [doi: 10.1057/palgrave.ivs.9500019]

57. Wilkinson L, Friendly M. The History of the Cluster Heat Map. Am Stat 2009;63(2):179-184. [doi: 10.1198/tas.2009.0033]

58. Hoeber O, Yang X. HotMap: supporting visual exploration of web search results. J Am Soc Inf Sci 2009 Jan;60(1):90-110. [doi: 10.1002/asi.20957]

Abbreviations

HPO: Human Phenotype Ontology
MEDLINE: Medical Literature Analysis and Retrieval System Online
MeSH: Medical Subject Header
NLM: National Library of Medicine
OVERT-MED: Ontology-Driven Visual Search and Triage Interface for MEDLINE

Edited by M Focsa; submitted 30.10.16; peer-reviewed by PCI Pang, A Benis; comments to author 26.11.16; accepted 03.01.17; published 02.02.17

Please cite as:
Demelo J, Parsons P, Sedig K
Ontology-Driven Search and Triage: Design of a Web-Based Visual Interface for MEDLINE
JMIR Med Inform 2017;5(1):e4
URL: http://medinform.jmir.org/2017/1/e4/
doi:10.2196/medinform.6918
PMID:28153818