

Supporting information retrieval from electronic health records: A report of University of Michigan's nine-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE)



David A. Hanauer^{a,b,*}, Qiaozhu Mei^{b,c}, James Law^d, Ritu Khanna^d, Kai Zheng^{b,e}

^a Department of Pediatrics, University of Michigan Medical School, Ann Arbor, MI, USA

^b School of Information, University of Michigan, Ann Arbor, MI, USA

^c Department of Electronic Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA

^d Medical Center Information Technology, University of Michigan Health System, Ann Arbor, MI, USA

^e School of Public Health, Department of Health Management and Policy, University of Michigan, Ann Arbor, MI, USA

ARTICLE INFO

Article history:

Received 22 January 2015

Revised 31 March 2015

Accepted 5 May 2015

Available online 13 May 2015

Keywords:

Electronic health records

(E05.318.308.940.968.625.500)

Search engine (L01.470.875)

Information storage and retrieval (L01.470)

ABSTRACT

Objective: This paper describes the University of Michigan's nine-year experience in developing and using a full-text search engine designed to facilitate information retrieval (IR) from narrative documents stored in electronic health records (EHRs). The system, called the Electronic Medical Record Search Engine (EMERSE), functions similar to Google but is equipped with special functionalities for handling challenges unique to retrieving information from medical text.

Materials and methods: Key features that distinguish EMERSE from general-purpose search engines are discussed, with an emphasis on functions crucial to (1) improving medical IR performance and (2) assuring search quality and results consistency regardless of users' medical background, stage of training, or level of technical expertise.

Results: Since its initial deployment, EMERSE has been enthusiastically embraced by clinicians, administrators, and clinical and translational researchers. To date, the system has been used in supporting more than 750 research projects yielding 80 peer-reviewed publications. In several evaluation studies, EMERSE demonstrated very high levels of sensitivity and specificity in addition to greatly improved chart review efficiency.

Discussion: Increased availability of electronic data in healthcare does not automatically warrant increased availability of information. The success of EMERSE at our institution illustrates that free-text EHR search engines can be a valuable tool to help practitioners and researchers retrieve information from EHRs more effectively and efficiently, enabling critical tasks such as patient case synthesis and research data abstraction.

Conclusion: EMERSE, available free of charge for academic use, represents a state-of-the-art medical IR tool with proven effectiveness and user acceptance.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Background and significance

In addition to improving patient care delivery, the widespread adoption of electronic health records (EHRs) in the U.S. has created unprecedented opportunities for increased access to clinical data, enabling multiple secondary use purposes such as quality assurance, population health management, and clinical and translational research. The broader use of clinical data for discovery,

* Corresponding author at: 5312 CC, SPC 5940, 1500 E Medical Center Drive, Ann Arbor, MI 48109-5940, USA. Tel.: +1 (734) 615 0599; fax: +1 (206) 338 4213.

E-mail address: hanauer@med.umich.edu (D.A. Hanauer).

surveillance, and improving care provides great potential to transform the U.S. healthcare system into a self-learning vehicle—or a “Learning Health System”—to advance our knowledge in a wide range of clinical and policy domains [1,2].

However, the benefits of electronically captured clinical data have yet to be fully realized for a number of reasons. Foremost is the continued popularity of free-text documentation in EHRs. While structured data at the time of entry is desirable, unstructured clinical documentation is likely to persist due to the need by clinicians to express their thoughts in a flexible manner and to preserve the complexity and nuances of each patient [3,4]. Recent studies have shown that clinicians often revert to

free-text entry even when coding options are provided [3,5–7], and that the free text is still needed for complex tasks such as clinical trial recruitment [8].

The challenges to extracting information locked in medical text should not be underestimated [9]. Factors contributing to this complexity include clinicians' frequent use of interchangeable terms, acronyms and abbreviations [10], as well as negation and hedge phrases [11–13]. Ambiguities may also arise due to a lack of standard grammar and punctuation usage [10] and the inherent difficulties for computer systems to process context-sensitive meanings [14], anaphora and coreferences [15], and temporal relationships [16]. Even different ways by which clinical notes were created (e.g., via dictation/transcription vs. typing) could result in distinct linguistic properties posing post-processing challenges [17]. Indeed, a paradox has been noted in the biomedical informatics literature that increased availability of electronic patient notes does not always lead to increased availability of information [18].

Automated data extraction methods, including natural language processing, hold great promise for transforming unstructured clinical notes into a structured, codified, and thus computable format [19]. However, the use of such tools is often associated with considerable upfront costs in software setup and in training the algorithms for optimal performance. Further, despite significant research advancements, the precision and recall of such tools are not yet up to par for meeting the requirements of many sophisticated chart abstraction tasks, and existing tools' lack of generalizability often necessitates customized solutions be built to the specific needs and data characteristics of each problem [20–23].

As such, search engines, or information retrieval (IR) systems more generally, offer an effective, versatile, and scalable solution that can augment the value of unstructured clinical data [24–29]. Search engines help human reviewers quickly pinpoint where information of interest is located, while leaving some difficult problems that computers are not yet capable of solving to human wisdom. The requirement for end user training is also minimized as healthcare practitioners and researchers are already familiar with how search engines work through their day-to-day interactions with general-purpose web search engines such as Google and literature search tools such as PubMed. This antecedent familiarity is important because in healthcare, clinicians, administrators, and researchers often lack time to attain mastery of informatics tools.

Surprisingly, despite a growing need for IR tools in healthcare settings for both operational and research purposes, very few successful implementations have been reported in the literature [26,27,30]. In this paper, we describe the University of Michigan's (UM) nine-year experience in developing and using a web-based full-text medical search engine designed to facilitate information retrieval from narrative EHR documents. The system, called the Electronic Medical Record Search Engine (EMERSE), has been used by numerous research groups in over 750 clinical and translational studies yielding 80 peer-reviewed publications to date (e.g., [31–35]). As part of the results validation process several studies explicitly examined the efficacy of EMERSE and concluded that the system was instrumental in ensuring the quality of chart review while significantly reducing manual efforts [33,36,37]. To the best of our knowledge, this is the first comprehensive description of an EHR search tool that has been used in a production setting by many real end users for multiple years and for a wide variety of clinical, operational, and research tasks. We believe our experience of designing, building, maintaining, and disseminating EMERSE will provide useful insights to researchers who have a similar need for an EHR search engine and will help potential users of such an EHR search engine to understand what types of problems that such tools can help solve.

In the following sections, we first describe the background and architecture of EMERSE, followed by a presentation of various

usage metrics and the results of user adoption and effectiveness evaluations. Note that the primary purpose of this paper is to provide a comprehensive description of the design and features of EMERSE, especially those that significantly deviate from traditional IR systems and those that are well received among end users of EMERSE, most of whom are clinicians, healthcare administrators, and clinical researchers. It should also be noted that there is no universally accepted benchmark, or 'gold standard,' with which to measure the performance of a system such as EMERSE that is designed to serve a wide range of purposes in a wide variety of clinical contexts. For example, when trying to identify a subset of potential study subjects among a pool of many possible candidates, precision may be most important. By contrast, when trying to identify all patients affected by a faulty pacemaker then recall may be most important. Therefore, our goal is not to evaluate the IR performance of the EMERSE and compare it to that of other search engines (e.g., PubMed) or algorithms developed for IR competitions (e.g., TREC), but rather to describe EMERSE in the context of the mostly empty landscape of EHR-specific IR tools. Such information is important to developing an enhanced understanding of how to achieve better penetration of IR systems such as EMERSE in everyday healthcare settings.

2. Methods and materials

2.1. History and current status

EMERSE has been operational since 2005 and has undergone multiple rounds of interface and architectural revisions based on end user feedback, usability testing, and the changing technology environment. The National Center for Advancing Translational Sciences, the National Library of Medicine (NLM), the UM Comprehensive Cancer Center, the Michigan Institute for Clinical and Health Research, and the UM Medical Center Information Technology department provided funding support for the software's continued development and evaluation.

EMERSE was originally designed to work with the University of Michigan Health System's (UMHS) legacy homegrown EHR system, CareWeb, deployed in 1998. In 2012, EMERSE was overhauled so it could integrate data from our newly implemented commercial EHR system, Epic (Epic Systems Corporation, Verona, WI), locally renamed MiChart. During the overhaul, significant efforts were made to ensure EMERSE is as much platform independent and vendor-neutral as possible. At present, users at UMHS can use EMERSE to search through 81.7 million clinical documents: 36.4 million from CareWeb, 10.6 million from Epic (MiChart), 10.4 million radiology reports, 23.2 million narrative pathology reports, and 1.2 million other genres of documents such as electroencephalography and pulmonary function studies.

EMERSE has also been adapted to work with VistA, the Veterans Affairs (VA)'s health IT architecture, and has been used at the VA Hospital in Ann Arbor, Michigan since 2006 to support various VA research initiatives [38–40]. This adapted version can be readily adopted by other VA facilities nationwide as they all share the same underlying IT infrastructure. EMERSE is available free of charge for academic use. Note that due to reasons such as local customization, not all features described in this paper are present in all versions of the software. Additional details, as well as a demonstration version of EMERSE, are available at <http://project-emerse.org>.

2.2. User interface design and software architecture

Fig. 1 exhibits the main workspace of EMERSE where users construct search queries and subsequently submit the queries to the

The screenshot shows the EMERSE main workspace. At the top, there are tabs for "Patients" (Animal bites (15 Patients)), "Dates" (All), and "Terms" (cat bite, dog bite, puncture wound, irrigate, irrigating, irrigates, irrigated, irrigation). A "Search" button is visible. Below this is a "Quick Search" section with a text input field and a "Set Search Terms" button. Underneath is a "Saved Search (Bundles)" section with a "Filter Bundles" input, a "New Bundle" button, and radio buttons for "Favorites (11)" and "All Available (45)". A table lists various search bundles with columns for Name, Description, Owner, Last Updated, Rating, Term Count, and Favorite.

Name	Description	Owner	Last Updated	Rating	Term Count	Favorite
Animal bites	Finding mentions of dog, cat, and other bites	David Hanauer	07/10/2014	★★★★★	8	<input checked="" type="checkbox"/>
Chest pain	Identifying symptoms of chest pain	David Hanauer	07/02/2014	★★★★★	16	<input checked="" type="checkbox"/>
Constipation	Meds that cause and meds that treat consti...	David Hanauer	05/15/2014	★★★★★	77	<input checked="" type="checkbox"/>
DRR - Stroke	For identifying strokes in patients with d...	David Hanauer	09/27/2012	★★★★★	17	<input checked="" type="checkbox"/>
Depression	Related to depression and medications to t...	David Hanauer	05/15/2014	★★★★★	32	<input checked="" type="checkbox"/>
GVHD bundle	GVHD in pediatric HSCT patients	David Hanauer	05/15/2014	★★★★★	13	<input checked="" type="checkbox"/>
HTN	Hypertension terms	David Hanauer	07/02/2014	★★★★★	36	<input checked="" type="checkbox"/>
Influenza-like illness	Signs and symptoms that could be related t...	David Hanauer	05/15/2014	★★★★★	27	<input checked="" type="checkbox"/>
Myocardial infarction	Labs, signs, and symptoms related to an MI	David Hanauer	05/15/2014	★★★★★	56	<input checked="" type="checkbox"/>
Otitis Media	Terms related to ear infections and treatment	David Hanauer	05/15/2014	★★★★★	3	<input checked="" type="checkbox"/>
Steroids	Indications for steroid treatment	David Hanauer	05/15/2014	★★★★★	14	<input checked="" type="checkbox"/>

Fig. 1. EMERSE main workspace. Screen capture of EMERSE showing where search terms can be entered. Search keywords can be quickly typed into the *Quick Search* text box. Available pre-saved collections of search terms (i.e., search term bundles) are listed further down on the same page. In this example, a search term bundle named "Animal bites" has been selected and the terms belonging to the bundle are shown at the top.

backend IR engine for processing. Search terms can be entered rapidly as a *Quick Search* or they can be activated from *Search Term Bundles* pre-stored in the system.

2.2.1. Quick Search

Similar to Google, the most common way of using EMERSE is to type keywords into a simple text entry box. Search terms may contain single words or multi-word phrases (e.g., "sick sinus syndrome"), wild cards (e.g., "hyperten*"), and other operators (e.g., ^ for case sensitivity). In earlier versions of EMERSE, advanced users could also write sophisticated search queries using regular expressions. This function was dropped during the 2012 overhaul due to lack of use.

Searches in EMERSE are case insensitive by default, but an option is provided allowing users to enforce the case-sensitivity, such as for distinguishing "FROM" (full range of motion) from the common word "from." Similarly, stop words are preserved in the document indices because many are legitimate acronyms of medical concepts, e.g., OR: operating room; IS: incentive spirometry; IT: intrathecal.

Exclusion criteria can be entered to instruct the system not to include certain words and phrases in the search. This feature has been utilized particularly in handling negations. For example, the UMHS Department of Ophthalmology developed a "search term

bundle" (see below) to look in surgeon notes for perioperative complications (Appendix A.1). The query contains only one search term, "complications," while excluding 51 phrases that unambiguously rule out the possibility of perioperative complications (e.g., "without any complications") or that mentioned complications in an irrelevant context (e.g., "diabetes with neurologic complications").

2.2.2. Search term bundles and collaborative search

EMERSE provides a special "collaborative search" mechanism that allows users to save their search queries as "search term bundles" which can be reused and shared with others. Examples include a bundle that contains 28 search terms enumerating common ways in which apathy or indifference may be described in clinician notes (Appendix A.2), and another that lists 70 concepts for identifying infections in hematopoietic stem cell transplant patients (Appendix A.3). This collaborative search feature was inspired by social information foraging and crowdsourcing techniques found on the Web that leverage users' collective wisdom to perform collaborative tasks such as IR. The resulting search term bundles not only provide a means for end users to preserve and collectively refine search knowledge, but also to ensure the consistent use of standardized sets of search terms by users. In prior work assessing adoption of this feature, we found that about half

of the searches performed in EMERSE had used pre-stored search term bundles, of which one-third utilized search knowledge shared by other users [41].

2.2.3. Handling of spelling errors and logic validation

Medical terminology contains many difficult-to-spell words (e.g., “ophthalmology”) which can be challenging even to seasoned clinicians. It is also not uncommon for misspellings to make their way into official patient records. These can be words incorrectly spelled such as “infectoin” instead of “infection,” or words that were spelled correctly (thus eluding detection by a spell checker) but were nevertheless incorrect, such as “prostrate” instead of “prostate”.

To address these issues, EMERSE incorporates a medical spelling checker to alert users about potentially misspelled words in their search queries. In addition, EMERSE offers an option for users to include potentially misspelled forms of the search terms in the search. The implementation of these features was based on phonetic matching and sequence comparison algorithms provided in open-source spell check APIs (Jazzy in earlier versions of EMERSE and Apache Lucene in the overhauled version) [42,43], and a customized dictionary containing about 6200 common spelling alternatives that we manually curated from the search logs of EMERSE over the years. The dictionary includes, for example, 24 misspelled

forms of the word “diarrhea”, e.g. “diarrhea” and “diarheea”. Besides spelling mistakes, EMERSE also inspects for common logic errors found in user-submitted search queries, e.g., same keywords appearing on both inclusion and exclusion lists.

2.2.4. Query recommendation

EMERSE users who are tasked with reviewing medical documents do not necessarily possess adequate clinical knowledge (e.g., they may be student research assistants). Through our observations of users and analyses of search logs, we also discovered that even clinicians with extensive clinical experience might have difficulty creating a set of search terms ‘minimally necessary’ to ensure reasonably inclusive search results [44]. For example, when looking for “myocardial infarction,” users often failed to include common synonyms such as “heart attack,” “cardiac arrest,” and its acronym “MI”.

To improve search quality and reduce user variation, significant effort was made to build a query expansion function to recommend alternative terms that users may consider adding to a query (Fig. 2). These alternative terms can be acronyms and synonyms of the keywords searched, as well as generic names of a commercial drug or vice versa. The knowledge base underlying this feature, currently consisting of about 78,000 terms representing approximately 16,000 concepts, was derived from multiple sources

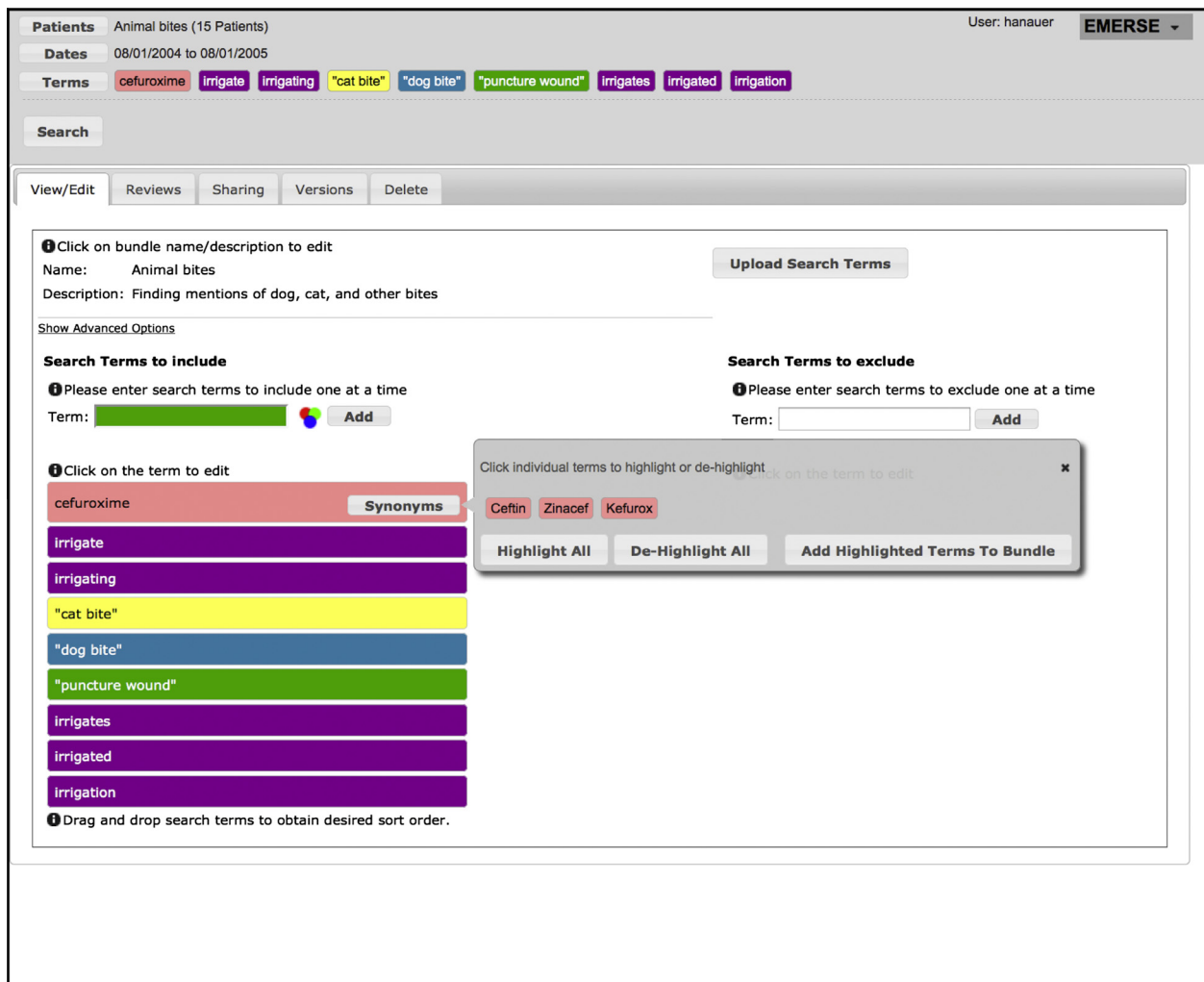


Fig. 2. Screen capture of EMERSE showing the editing of a search terms bundle. In the example, the term “cefuroxime” was added by the user and the system provided additional suggestions. To aid in recognizing terms in the search results, the bundles feature will highlight synonyms and related concepts using the same color, which can be overridden by users. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

including medical dictionaries, drug lexicons, and an empirical synonym set manually curated from the search logs of EMERSE and from the pre-stored search term bundles. The query recommendation feature is available with the 'Quick Search' option as well as with the 'Bundles' option.

Funded by the NLM, we also developed an experimental extension to EMERSE to leverage the nomenclatures included in the Unified Medical Language System® (UMLS®) Metathesaurus for more comprehensive query expansion. This experimental extension also incorporates MetaMap, NLM's named entity recognition engine [45], to enable the use of additional text features such as term frequency, inverse document frequency, and document length penalization to improve the relevance of document ranking of the search results that EMERSE returns.

2.2.5. Multilevel data views with visual cues

EMERSE presents search results through multilevel data views and uses visual cues to help users quickly scan through returned documents. The data view at the highest level, the *Overview*, provides users a succinct summary of patients (rows) and document sources (columns) where the "hits" of a search are found (Fig. 3). Cells in the *Overview* display a color gradation along with both the number of relevant documents found and the total number of documents for the patient, suggesting the 'intensity' of the hits. Clicking on a cell in the *Overview* takes users to the *Summaries* view

where documents are presented in reverse chronological order with text snippets revealing the context in which the search terms appear (Fig. 4). Clicking on a snippet will then reveal the full document in the *Documents* view.

There are two important design aspects of EMERSE that deviate from a general-purpose search engine with respect to these views. The first is that documents are grouped and displayed by patient, and among each patient they are grouped and displayed by source system (e.g., pathology, radiology, etc.). The second is that all documents for a patient are retrieved and made available to the user, including those without a hit of the query words. This is visible in the top row of the results shown in Fig. 4. On that row, no visible snippets are displayed, demonstrating that the document does not contain any of the search terms used. Retaining this document in the *Documents* view is still important in many common scenarios of EHR search because even though such documents do not contain the exact search terms, they may still be relevant to the particular task of the user (e.g., patient screening).

Matched keywords found in returned documents (or in text snippets displayed in the document summary view) are automatically highlighted using a palette of 18 system-assigned colors to ensure that terms stand out from their surroundings (Fig. 2). For search terms included in pre-stored bundles, users have additional controls over the choice of colors; for example, they can override system-assigned colors to allow logical groupings by color of

MRN	Patient Name	MIChart	CareWeb	Radiology	Pathology	Other
1234567890	Smith, James					
3141592654	Johnson, Mary					
0723197300	Williams, Elizabeth					
6022141690	Brown, Robert					
4815162342	Miller, Jennifer					
4325200327	Garcia, David		1 of 13			
0914197000	Martinez, Linda					
0987654321	Anderson, Susan					
0122320100	Moore, Charles					
0070920070	Lee, Daniel					
0114201300	White, Jessica					
6673848000	Gonzalez, Sarah					
0299792458	Harris, Emily			10 of 12		
6626069570	Walker, Ashley					
2718281828	Green, Carol		1 of 15			

Fig. 3. The Overview screen, presenting the search results for eight terms related to animal bites and wound cleaning. Each row on the Overview screen represents a patient and each column represents a document source. All patient identifiers have been replaced with realistic substitutes.

Summary	Case Date	Document Type	Service	Department	Read
	08/25/2004	PHONE NOTE	EMED	EAA	N
...Chief Complaint: Cat bite . History of Present Illness:... ...condition. Diagnosis: Cat bite	08/26/2004	ED NOTE	AER	ER	N
...with cellulitis after a dog bite . She reports that yesterday... ...about 36 hours out from a dog bite with a significant cellulitis... ...Final Impression: 1. Dog bite . 2. Cellulitis. Disposition:...	08/26/2004	ED NOTE	AER	ER	N
...Reason for Consultation: Dog bite to left hand and right... ... history of infected cat bite approximately 4 months... ...finger. These areas were irrigated and packed in the emergency...	08/27/2004	CONSULT-ED	SPLA	SUR	Y
...PREOPERATIVE DIAGNOSIS: Dog bite , left hand, with cellulitis... ...POSTOPERATIVE DIAGNOSIS: Dog bite , left hand, with cellulitis... ...woman with a history of a dog bite to the left hand. She... ...wounds were copiously irrigated with sterile saline....	09/05/2004	OPER-REPORT	SPLA	SUR	N
...Cellulitis, left hand after dog bite . COMORBIDITIES: None... ...as opening of the puncture wound tract in the long... ...wounds were copiously irrigated , and the patient was...	09/23/2004	D/C SUMMARY	SPLA	SUR	Y
...Briefly, she is status post cat bite to the left hand on January...	09/27/2004	LET/NOTE-RV	SPLA	SUR	N
...woman who sustained a dog bite to the left hand on January...	09/28/2004	LET/NOTE-RV	SPLA	SUR	N
...of her right hand puncture wound . They gave the patient... ...drainage from the puncture wound . Past Medical History:... ...approximately 5-mm puncture wound that entered in the... ...over the site of the puncture wound . Strength was intact... ...who presents with a puncture wound to her right hand... ...Diagnosis: Right hand puncture wound with cellulitis. ...	11/08/2004	ED NOTE	AER	ER	N
...recheck for a right hand puncture wound that occurred approximately... ...hand puncture wound . The hand puncture wound looked... ...right hand has the puncture wound just about 2 cm proximal... ...Recheck of right hand puncture wound . I saw & evaluated...	12/04/2004	ED NOTE	AER	ER	N
...complications from a dog bite . She has been off her...	12/06/2004	LET/NOTE-RV	EMED	EAA	N

Fig. 4. The document Summaries view where the rows represent each of the documents belonging to a specific patient. Documents that do not contain any of the search terms are also included for viewing if needed.

similar medical concepts (e.g., all narcotic medications in green and all stool softeners in orange). This feature makes rapid visual scanning of search results easier, especially when many search terms are involved. The use of distinct colors for different concepts is not commonly supported by general-purpose search engines, but this feature has been highly appreciated by our users.

2.2.6. Patient lists and date range

EMERSE allows users to supply a predefined list of patients to limit the scope of search which can be, for example, a cohort of patients that has passed preliminary trial eligibility screening. Such lists may be prepared *ad hoc* (e.g., of patients currently staying in an inpatient unit) or systematically identified through claims data, disease-specific patient registries, or research informatics tools such as the i2b2 Workbench [46]. We have developed a plug-in for the i2b2 Workbench so that users can directly transfer patient cohorts identified in i2b2 to EMERSE as predefined patient lists ready for searching. Searches in EMERSE can also be bounded with a date range. This feature is used by the UMHS Infection Control and Epidemiology Department to identify cases of post-operative surgical site infections in one-month blocks.

2.2.7. Security

EMERSE is hosted behind the UMHS firewall in a computing environment certified for storing protected health information.

Access to EMERSE is limited to authorized personnel who have patient data access privileges or, among those using it for research, have provided evidence of training in responsible research practices and proof of valid institutional review board approvals, including demonstration of a need to review identifiable patient information. At each login users must complete a brief attestation form to document their intent of use. Audit trail logs are kept for each use session. Searching clinical documents for clinical or operational purposes generally does not warrant removal of identifiers. For research, de-identification would be desirable in certain settings, but is not currently mandated by our health system. Nevertheless, we are exploring de-identification of the entire document corpus to reduce privacy concerns.

2.2.8. Technical implementation

The technical architecture of the most recent version of EMERSE is illustrated in Fig. 5. The system uses a variety of open-source components including Apache Lucene, Spring, and Hibernate, with a code base written in Java. Free-text documents are indexed nightly using Apache Solr without additional processing. Document repositories are recommended for indexing but are not needed at EMERSE runtime. Further, while our institution utilizes Epic as its EHR, EMERSE was built to be vendor-neutral and contains no dependencies on the EHR itself. Additional implementation details can be found in our online Technical Manual

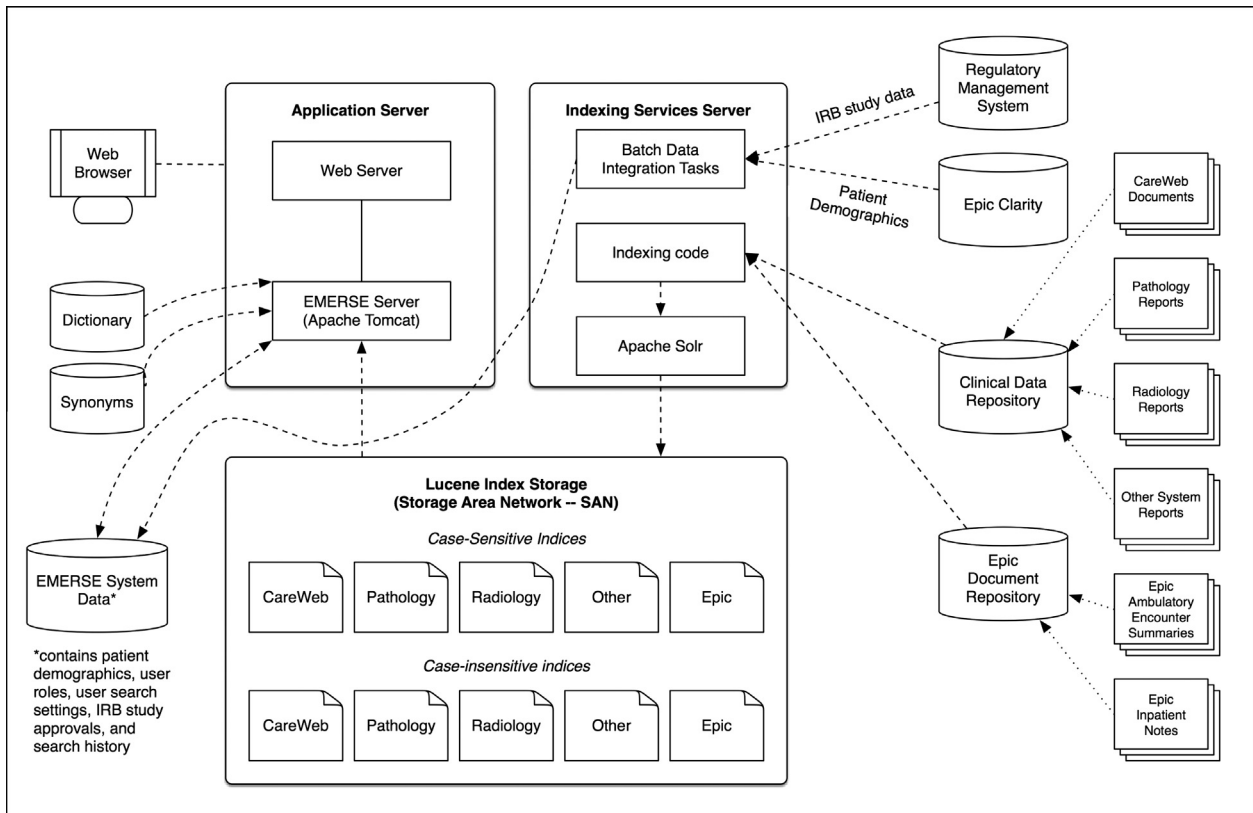


Fig. 5. The technical architecture of EMERSE. Data are captured from the sources of EHR data (right side) mainly via HL7 interfaces and added to locally-maintained document repositories separate from the EHR. The Indexing services server routinely updates the Lucene-based indices used by the main search application from these databases. Note that while having all documents indexed in both a case-sensitive and case-insensitive index increases storage requirements, this approach allows for faster response time for most search scenarios.

(<http://project-emerse.org/software.html>), including directions on how to configure EMERSE to work with local data sources.

3. Results

The most revealing result in evaluating a practical informatics tool such as EMERSE is perhaps whether people use the system in their everyday work. In this section, we present usage statistics of EMERSE, typical use scenarios, and results of several evaluation studies that assessed the IR performance and end user satisfaction of the system.

3.1. Usage statistics and typical use scenarios

As of March 2015, EMERSE has had 1,137 registered users who represent nearly all clinical and operational departments at UMHS covering a majority of medical specialties and professions. These users have collectively logged in more than 138,000 times. EMERSE currently stores 961 search term bundles created by users, each on average containing 32 search terms (median 18; range 1–774).

Several UMHS operational departments have incorporated EMERSE into their job routine in data synthesis and reporting. For example, the cancer registrars at our cancer center use EMERSE to perform data abstraction of genetic and biomarker testing results for submission to our tumor registry. Coding teams in the Department of Health Information Management use the system as a computer-assisted coding tool to improve the efficiency in evaluating and managing inpatient billing code assignments as well as for regulatory compliance reviews. EMERSE has also supported the coding team in improving hospital reimbursements by

aiding in the identification of supporting information (e.g., proof of pneumonia on admission) for reimbursements that were previously rejected by insurers, and for which they lacked the manpower to perform manual chart reviews. Professionals in the Department of Infection Control and Epidemiology use EMERSE to monitor the incidence of surgical site infections, a process that previously required labor-intensive manual chart auditing [47].

In the clinical setting, EMERSE has enabled providers to answer questions such as “for this patient in clinic today, what medication worked well for her severe migraine headache several years ago,” without having to read every note. EMERSE has also been used often in obtaining ‘prior authorization’ clearance for medications initially denied by insurance companies. With the system, clinic staff can rapidly identify supporting evidence as to when and why cheaper, more ‘preferable’ medications have failed for a patient.

Researchers at UMHS use EMERSE to perform a variety of chart abstraction tasks ranging from cohort identification and eligibility determination to augmenting phenotypic data for laboratory-based translational research linking biomarkers to patient outcomes [35,48–50]. The research studies that EMERSE has supported span many topic areas in health sciences [35,51,52]. Several papers explicitly acknowledged the enabling role that EMERSE played in the studies which allowed the investigators to “systematically identify data within the electronic medical record” [33], and helped to accomplish “a standardized, accurate, and reproducible electronic medical record review” [36,37]. Table 1 exhibits 14 peer-reviewed publications appearing in 2014 alone that were supported by the use of EMERSE. A full list of the publications supported by EMERSE to date, of which we are aware, is provided in Appendix B.

Table 1

Peer-reviewed publications supported by EMERSE, 2014 only.

Publication title	The way EMERSE was used	Journal	PubMed identifier
Anticoagulant complications in Facial Plastic and Reconstructive Surgery	Data abstraction from operative and other notes for wound healing, infection, bleeding, return to the operating room, and other factors	JAMA Facial Plast Surg	25541679
Proton pump inhibitors and histamine 2 blockers are associated with improved overall survival in patients with head and neck squamous carcinoma	To identify mentions of medications taken by the patients in the study	Cancer Prev Res	25468899
Transcutaneous biopsy of adrenocortical carcinoma is rarely helpful in diagnosis, potentially harmful, but does not affect patient outcome	To obtain data about staging, reason for biopsy, number of biopsies, institution performing the biopsy, complications, and other factors	Eur J Endocrinol	24836548
Engraftment syndrome after allogeneic hematopoietic cell transplantation predicts poor outcomes	To identify patients with engraftment syndrome using a set of 82 search terms	Biol Blood Marrow Transplant	24892262
Osteochondromas after radiation for pediatric malignancies: a role for expanded counseling for skeletal side effects	To identify patient who developed osteochondromas after hematopoietic stem cell transplant and total body irradiation	J Pediatr Orthop	23965908
Unbiased identification of patients with disorders of sex development	Chart reviews to locate features described in the notes that helped identify patients with disorders of sex development	PLoS One	25268640
Changes in characteristics of hepatitis C patients seen in a liver center in the United States during the last decade	Medical records review which included treatment status and clinical status at the time of presentation	J Viral Hepat	25311830
Symptomatic subcapsular and perinephric hematoma following ureteroscopy lithotripsy for renal calculi	Chart review of clinical notes to identify patients whose surgery was complicated by a symptomatic hematoma	J Endourol	25025758
Comparison of Second-Echelon treatments for Ménière's disease	Data abstraction for and identification of patients with Ménière's disease	JAMA Otolaryngol Head Neck Surg	25057891
A combined paging alert and web-based instrument alters clinician behavior and shortens hospital length of stay in Acute Pancreatitis	Finding data of interest for chart reviews	Am J Gastroenterol	24594946
Adjuvant therapies, patient and tumor characteristics associated with survival of adult patients with adrenocortical carcinoma	Chart reviews for a variety of clinical parameters including stage, pathology data, evaluation of hormone secretion, surgical approach, and treatment modalities	J Clin Endocrinol Metab	24302750
Population-based incidence and prevalence of systemic lupus erythematosus: The Michigan Lupus Epidemiology & Surveillance (MILES) Program	Data abstraction for a variety of clinical characteristics for patients with systemic lupus erythematosus	Arthritis Rheum	24504809
The association between race and gender, treatment attitudes, and antidepressant treatment adherence	Eligibility determination based on chart review by considering factors such as clinically significant depression and having a new recommendation for antidepressant treatment	Int J Geriatr Psychiatry	23801324
Aneurysms in abdominal organ transplant recipients	To identify patients with mention of an arterial aneurysm in the clinical documentation among patients undergoing liver or kidney transplants over an 11 year period	J Vasc Surg	24246534

3.2. Results of evaluation of IR performance

We evaluated the IR performance of EMERSE in the context of conducting several clinical or informatics studies. For example, in a psychiatric trial [53], we assessed the efficiency of eligibility screening with or without the system and showed that the team assisted by EMERSE achieved significant time savings while maintaining accuracy compared to the team doing manual chart review. In another study aiming to improve the process of data submission to the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP), we demonstrated that automated data extraction procedures powered by EMERSE attained high sensitivity (100.0% and 92.8%, respectively) and high specificity (93.0% and 95.9%, respectively) in identifying postoperative complications, which compared favorably with existing ACS NSQIP datasets manually prepared by clinical surgical nurses [54]. Finally, in collaboration with pediatric cardiologists at UMHS, we compared the performance of using EMERSE vs. three other specialty surgical registries to identify a rare tachyarrhythmia associated with congenital heart surgery. EMERSE was found to be the best-performing method, yielding the highest sensitivity (96.9%) and had comparable performance on other evaluation dimensions [55].

3.3. Results of end user satisfaction survey

Periodic satisfaction surveys have been conducted as part of an ongoing effort to collect users' feedback regarding the usefulness and usability of EMERSE. The most recent survey received

Table 2

End user satisfaction survey results.

	Does not meet (%)	Meets or exceeds (%)	Does not apply (%)
1. Enables effective searching	0.4	92.6	7.0
2. Solves my problem or facilitates the tasks I face	1.3	88.7	10.0
3. Saves overall time and effort	0.4	95.3	4.3
4. Allows me to get accurate answers	0.9	88.9	10.2
5. Helps me find data I might have otherwise missed or overlooked	0.4	93.1	6.5
6. Has expanded my ability to conduct chart reviews to areas previously impossible with manual review	1.8	78.8	19.4

responses from 297 users (60.5% of all active users at the time). According to the survey, the three largest user groups of EMERSE were clinicians with research responsibilities (22.2%), data analysts/managers (21.5%), and full-time researchers (14.4%). Over a quarter of the respondents were specialized in hematology and oncology (26.7%) followed by pediatrics (17.5%), general medicine (8.7%), and general surgery (8.3%). Among the respondents there were also 17 (5.7%) undergraduate students, 17 (5.7%) non-medical graduate students, and 13 (4.4%) medical students approved to participate in clinical research as student research assistants.

The survey questionnaire solicited end users' options about EMERSE such as to what extent the system facilitates data extraction tasks ("solves my problem or facilitates the tasks I face") and

how use of the system compares to manual chart review processes (“has expanded my ability to conduct chart reviews to areas previously impossible with manual review”). The results are shown in Table 2. Across all items, EMERSE users reported a high level of satisfaction with the system and the vast majority believed the system contributed to improved time efficiency and helped them find data that they “might have otherwise missed or overlooked”.

4. Discussion

Healthcare practitioners and researchers continue to seek efficient and effective informatics tools to support their needs for retrieving information from narrative patient records. EHR search engines, working similar to Google but equipped with features to accommodate the unique characteristics of medical text, provide a viable solution for meeting such needs. The success of EMERSE at our institution demonstrates the potential of using EHR search engines in clinical, operational, and research settings, enabling critical tasks such as patient case synthesis and research data abstraction.

However, only a handful of studies in the biomedical informatics literature have discussed the design and use of EHR search engines (e.g., Harvard’s Queriable Patient Inference Dossier [QPID] system and Columbia’s CSearch system) [24,26–30,56,57]; even fewer have rigorously evaluated the systems’ usability and IR performance, or reported usage metrics. There remains a paucity of knowledge as to how well the existing systems have been used in practice and how their design and implementation can be further improved. We believe this paper provides valuable insights into addressing this knowledge gap.

We learned several important lessons through interacting with end users of EMERSE and incorporating their feedback to continuously improve the system. First, we found that users in healthcare desire simple tools, and are willing to sacrifice efficiency for less learning and more cognitive comfort: some spent hours scrutinizing ‘imperfect’ search results returned by rudimentary queries, rather than trying to use the advanced search features readily available to help them improve the results [44]. For example, the early generations of EMERSE provided regular expressions support allowing users to construct sophisticated search queries that could yield desired results with high precision. An analysis of the search logs, however, revealed that this feature was minimally utilized [44]. Second, we found that the quality of search queries submitted by end users is generally poor, particularly in failing to consider essential alternative phrasings [44]. Several new functions, such as computer-aided query recommendation, were introduced as a result to assist users in constructing high-quality queries.

Third, users of EMERSE embraced the collaborative search feature more than anticipated. Users were not only enthusiastic in adopting and using search term bundles shared by others, but also in creating bundles and making them widely available. This search knowledge sharing is also taking place beyond EMERSE. Several studies enabled by the system published their search terms used to ensure other researchers could replicate the results [36,58–64]. As more and more studies now leverage EHR data for clinical and translational research, the development of a central repository for search terms used in extracting data from EHRs could serve as a useful and sharable knowledge base.

Lastly, because the goal of EHR search engines is to help users retrieve relevant documents rather than providing a precise answer to the question at hand, the usability of such systems, particularly in organizing the research results returned, is critical to their success. An EHR search engine with a well-designed user interface could thus be even more effective than a system with superior IR performance but with a sub-optimal interface. In EMERSE, we implemented many visual and cognitive aids to help

users iteratively explore search results, identify the text of interest, and read the information in its surrounding context to better interpret the meaning. Based on user feedback, we believe that the usability of the system has played a substantial role in growing our user base and retaining existing users. Nevertheless, building IR functionalities into EHRs is complex, and there remains much to do to better understand the varied needs of the many individuals and groups who wish to extract information from narrative clinical documents.

5. Conclusion

This paper reports University of Michigan’s nine-year experience in developing and using a full-text search engine, EMERSE, designed to facilitate retrieval of information from narrative clinician notes stored in EHRs. Based on substantial adoption for a wide range of use scenarios, EMERSE has proven to be a valuable addition to the clinical, operational, and research enterprise at our institution. For other institutions interested in adopting EMERSE, the software is available at no cost for academic use.

Contributorship statement

David Hanauer contributed to the conception and design of the production system as well as the experimental extension to EMERSE, acquisition, analysis and interpretation of the data, and drafting the manuscript.

James Law contributed to the conception, design, and coding of the production system, analysis and interpretation of the data, and drafting the manuscript.

Ritu Khanna contributed to the conception, design, and coding of the production system, analysis and interpretation of the data, and drafting the manuscript.

Qiaozhu Mei contributed to conception and design of the experimental extension to EMERSE, the analysis and interpretation of the data, and drafting the manuscript.

Kai Zheng contributed to conception and design of the experimental extension to EMERSE, the analysis and interpretation of the data, and drafting the manuscript.

All authors approved the final version to be published and agree to be accountable for all aspects of the work.

Conflict of Interest statement

EMERSE was developed at the University of Michigan, and is currently being offered at no cost for academic use. For commercial use a non-academic license is being offered by the University of Michigan for a small fee. As inventor of EMERSE, David Hanauer is entitled to a portion of the royalties received by the University of Michigan.

Acknowledgments

This work was supported in part by the Informatics Core of the University of Michigan Cancer Center in addition to support received from the National Institutes of Health through the UMCC Support Grant (CA46592), the Michigan Institute for Clinical and Health Research through the Clinical Translational Science Award (UL1RR024986), and the National Library of Medicine (HHSN27620100032C). We would also like to acknowledge the support of Stephen Gendler and Kurt Richardson from the UMHS Medical Center Information Technology Team, Frank Manion from the Comprehensive Cancer Center, and Thomas Shanley from the CTSA-supported Michigan Institute for Clinical and Health Research.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jbi.2015.05.003>.

References

- [1] C. Friedman, M. Rigby, Conceptualising and creating a global learning health system, *Int. J. Med. Inform.* 82 (4) (2013) e63–e71.
- [2] C.P. Friedman, A.K. Wong, D. Blumenthal, Achieving a nationwide learning health system, *Sci. Transl. Med.* 2 (57) (2010) 57cm29.
- [3] E. Ford, A. Nicholson, R. Koeling, et al., Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: what information is hidden in free text?, *BMC Med Res. Methodol.* 13 (1) (2013) 105.
- [4] S.T. Rosenbloom, J.C. Denny, H. Xu, et al., Data from clinical notes: a perspective on the tension between structure and flexible documentation, *J. Am. Med. Inform. Assoc.* 18 (2) (2011) 181–186.
- [5] S. Madani, E. Shultz, An analysis of free text entry within a structured data entry system, *AMIA Annu. Symp. Proc.* 947 (2012).
- [6] K. Zheng, D.A. Hanauer, R. Padman, et al., Handling anticipated exceptions in clinical care: investigating clinician use of ‘exit strategies’ in an electronic health records system, *J. Am. Med. Inform. Assoc.* 18 (6) (2011) 883–889.
- [7] L. Zhou, L.M. Mahoney, A. Shakurova, et al., How many medication orders are entered through free-text in EHRs?—a study on hypoglycemic agents, *AMIA Annu. Symp. Proc.* 2012 (2012) 1079–1088.
- [8] P. Raghavan, J.L. Chen, E. Fosler-Lussier, et al., How essential are unstructured clinical narratives and information fusion to clinical trial recruitment?, *AMIA Joint Summits Transl. Sci. Proc. AMIA Summit Transl. Sci.* 2014 (2014) 218–223.
- [9] T. Edinger, A.M. Cohen, S. Bedrick, et al., Barriers to retrieving patient information from electronic health record data: failure analysis from the TREC medical records track, *AMIA Annu. Symp. Proc.* 2012 (2012) 180–188.
- [10] R.C. Barrows Jr, M. Busuioc, C. Friedman, Limited parsing of notational text visit notes: ad-hoc vs. NLP approaches, *Proc. AMIA Symp.* (2000) 51–55.
- [11] S. Agarwal, H. Yu, Biomedical negation scope detection with conditional random fields, *J. Am. Med. Inform. Assoc.* 17 (6) (2010) 696–701.
- [12] D.A. Hanauer, Y. Liu, Q. Mei, et al., Hedging their bets: the use of uncertainty terms in clinical documents and its potential implications when sharing the documents with patients, *AMIA Annu. Symp. Proc.* 2012 (2012) 321–330.
- [13] K.J. Mitchell, M.J. Becich, J.J. Berman, et al., Implementation and evaluation of a negation tagger in a pipeline-based system for information extract from pathology reports, *Stud. Health Technol. Inform.* 107 (Pt 1) (2004) 663–667.
- [14] V. Jagannathan, C.J. Mullett, J.G. Arbogast, et al., Assessment of commercial NLP engines for medication information extraction from dictated clinical notes, *Int. J. Med. Inform.* 78 (4) (2009) 284–291.
- [15] W.W. Chapman, G.K. Savova, J. Zheng, et al., Anaphoric reference in clinical reports: characteristics of an annotated corpus, *J. Biomed. Inform.* 45 (3) (2012) 507–521.
- [16] B. Tang, Y. Wu, M. Jiang, et al., A hybrid system for temporal information extraction from clinical text, *J. Am. Med. Inform. Assoc.* 20 (5) (2013) 828–835.
- [17] K. Zheng, Q. Mei, L. Yang, et al., Voice-dictated versus typed-in clinician notes: linguistic properties and the potential implications on natural language processing, *AMIA Annu. Symp. Proc.* 2011 (2011) 1630–1638.
- [18] T. Christensen, A. Grimsmo, Instant availability of patient records, but diminished availability of patient information: a multi-method study of GP’s use of electronic patient records, *BMC Med. Inform. Decis. Mak.* 8 (2008) 12.
- [19] A.K. Jha, The promise of electronic records: around the corner or down the road?, *JAMA* 306 (8) (2011) 880–881.
- [20] C. Friedman, T.C. Rindfleisch, M. Corn, Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine, *J. Biomed. Inform.* 46 (5) (2013) 765–773.
- [21] P.M. Nadkarni, L. Ohno-Machado, W.W. Chapman, Natural language processing: an introduction, *J. Am. Med. Inform. Assoc.* 18 (5) (2011) 544–551.
- [22] M.H. Stanfill, M. Williams, S.H. Fenton, et al., A systematic literature review of automated clinical coding and classification systems, *J. Am. Med. Inform. Assoc.* 17 (6) (2010) 646–651.
- [23] Q.T. Zeng, S. Goryachev, S. Weiss, et al., Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system, *BMC Med. Inform. Decis. Mak.* 6 (2006) 30.
- [24] P. Biron, M.H. Metzger, C. Pezet, et al., An information retrieval system for computerized patient records in the context of a daily hospital practice: the example of the Leon Berard Cancer Center (France), *Appl. Clin. Inform.* 5 (1) (2014) 191–205.
- [25] M. Gardner, Information retrieval for patient care, *BMJ* 314 (7085) (1997) 950–953.
- [26] W. Gregg, J. Jirjis, N.M. Lorenzi, et al., StarTracker: an integrated, web-based clinical search engine, *AMIA Annu. Symp. Proc.* (2003) 855.
- [27] K. Natarajan, D. Stein, S. Jain, et al., An analysis of clinical queries in an electronic health record search utility, *Int. J. Med. Inform.* 79 (7) (2010) 515–522.
- [28] A.A. Tawfik, K.M. Kochendorfer, D. Saparova, et al., “I Don’t Have Time to Dig Back Through This”: the role of semantic search in supporting physician information seeking in an electronic health record, *Perform. Improvement Quart.* 26 (4) (2014) 75–91.
- [29] M. Zalis, M. Harris, Advanced search of the electronic medical record: augmenting safety and efficiency in radiology, *J. Am. Coll. Radiol.* 7 (8) (2010) 625–633.
- [30] H.J. Lowe, T.A. Ferris, P.M. Hernandez, et al., STRIDE – an integrated standards-based translational research informatics platform, *AMIA Annu. Symp. Proc.* 2009 (2009) 391–395.
- [31] D.W. Blayney, K. McNiff, D. Hanauer, et al., Implementation of the quality oncology practice initiative at a university comprehensive cancer center, *J. Clin. Oncol.* 27 (23) (2009) 3802–3807.
- [32] T.J. Braley, B.M. Segal, R.D. Chervin, Sleep-disordered breathing in multiple sclerosis, *Neurology* 79 (9) (2012) 929–936.
- [33] M.J. DiMagno, J.P. Spaete, D.D. Ballard, et al., Risk models for post-endoscopic retrograde cholangiopancreatography pancreatitis (PEP): smoking and chronic liver disease are predictors of protection against PEP, *Pancreas* 42 (6) (2013) 996–1003.
- [34] D.A. Hanauer, N. Ramakrishnan, L.S. Seyfried, Describing the relationship between cat bites and human depression using data from an electronic health record, *PLoS One* 8 (8) (2013) e70585.
- [35] S. Paczesny, T.M. Braun, J.E. Levine, et al., Elafin is a biomarker of graft-versus-host disease of the skin, *Sci. Transl. Med.* 2 (13) (2010) 13ra2.
- [36] A.T. DeBenedet, T.E. Raghunathan, J.J. Wing, et al., Alcohol use and cigarette smoking as risk factors for post-endoscopic retrograde cholangiopancreatography pancreatitis, *Clin. Gastroenterol. Hepatol.* 7 (3) (2009) 353–8e4.
- [37] A.T. DeBenedet, S.D. Saini, M. Takami, et al., Do clinical characteristics predict the presence of small bowel angioectasias on capsule endoscopy?, *Dig Dis. Sci.* 56 (6) (2011) 1776–1781.
- [38] H.M. Kim, E.G. Smith, D. Ganoczy, et al., Predictors of suicide in patient charts among patients with depression in the Veterans Health Administration health system: importance of prescription drug and alcohol abuse, *J. Clin. Psychiatry* 73 (10) (2012) e1269–e1275.
- [39] H.M. Kim, E.G. Smith, C.M. Stano, et al., Validation of key behaviourally based mental health diagnoses in administrative data: suicide attempt, alcohol abuse, illicit drug abuse and tobacco use, *BMC Health Serv. Res.* 12 (2012) 18.
- [40] E.G. Smith, H.M. Kim, D. Ganoczy, et al., Suicide risk assessment received prior to suicide death by Veterans Health Administration patients with a history of depression, *J. Clin. Psychiatry* 74 (3) (2013) 226–232.
- [41] K. Zheng, Q. Mei, D.A. Hanauer, Collaborative search in electronic health records, *J. Am. Med. Inform. Assoc.* 18 (3) (2011) 282–291.
- [42] The Java Open Source Spell Checker. Secondary The Java Open Source Spell Checker. <<http://jazzy.sourceforge.net/>>.
- [43] SpellChecker. Secondary SpellChecker. <<http://wiki.apache.org/lucene-java/SpellChecker>>.
- [44] L. Yang, Q. Mei, K. Zheng, et al., Query log analysis of an electronic health record search engine, *AMIA Annu. Symp. Proc.* 2011 (2011) 915–924.
- [45] A.R. Aronson, F.M. Lang, An overview of MetaMap: historical perspective and recent advances, *J. Am. Med. Inform. Assoc.* 17 (3) (2010) 229–236.
- [46] S.N. Murphy, G. Weber, M. Mendis, et al., Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), *J. Am. Med. Inform. Assoc.* 17 (2) (2010) 124–130.
- [47] C. Friedman, L.K. Sturm, C. Chenoweth, Electronic chart review as an aid to postdischarge surgical site surveillance. increased case finding, *Am. J. Infect. Control* 29 (5) (2001) 329–332.
- [48] J.L. Ferrara, A.C. Harris, J.K. Greenson, et al., Regenerating islet-derived 3-alpha is a biomarker of gastrointestinal graft-versus-host disease, *Blood* 118 (25) (2011) 6702–6708.
- [49] A.C. Harris, J.L. Ferrara, T.M. Braun, et al., Plasma biomarkers of lower gastrointestinal and liver acute GVHD, *Blood* 119 (12) (2012) 2960–2963.
- [50] S. Paczesny, O.I. Krijanovski, T.M. Braun, et al., A biomarker panel for acute graft-versus-host disease, *Blood* 113 (2) (2009) 273–278.
- [51] M. Gao, J.J. Gemmete, N. Chaudhary, et al., A comparison of particulate and onyx embolization in preoperative devascularization of juvenile nasopharyngeal angiofibromas, *Neuroradiology* 55 (9) (2013) 1089–1096.
- [52] K.M. Jensen, L.C. Taylor, M.M. Davis, Primary care for adults with down syndrome: adherence to preventive healthcare recommendations, *J. Intellect. Disabil. Res.* 57 (5) (2013) 409–421.
- [53] L. Seyfried, D.A. Hanauer, D. Nease, et al., Enhanced identification of eligibility for depression research using an electronic medical record search engine, *Int. J. Med. Inform.* 78 (12) (2009) e13–e18.
- [54] D.A. Hanauer, M.J. Englesbe, J.A. Cowan Jr, et al., Informatics and the American College of Surgeons National Surgical Quality Improvement Program: automated processes could replace manual record review, *J. Am. Coll. Surg.* 208 (1) (2009) 37–41.
- [55] J.D. Zampi, J.E. Donohue, J.R. Charpie, et al., Retrospective database research in pediatric cardiology and congenital heart surgery: an illustrative example of limitations and possible solutions, *World J. Pediatr. Congenit. Heart Surg.* 3 (3) (2012) 283–287.
- [56] E.J. Campbell, A. Krishnaraj, M. Harris, et al., Automated before-procedure electronic health record screening to assess appropriateness for GI endoscopy and sedation, *Gastrointest. Endosc.* 76 (4) (2012) 786–792.
- [57] E.L. Scheufele, D. Housman, J.S. Wasser, et al., i2b2 and keyword search of narrative clinical text, *Annu. Symp. Proc.* 2011 (1950).

- [58] R. Asmar, J.L. Beebe-Dimmer, K. Korgavkar, et al., Hypertension, obesity and prostate cancer biochemical recurrence after radical prostatectomy, *Cancer Prostatic Dis.* 16 (1) (2013) 62–66.
- [59] N. Ghaziuddin, D. Dhossche, K. Marcotte, Retrospective chart review of catatonia in child and adolescent psychiatric patients, *Acta Psychiatr. Scand.* 125 (1) (2012) 33–38.
- [60] K.M. Jensen, C.R. Cooke, M.M. Davis, Fidelity of administrative data when researching down syndrome, *Med. Care* (2013).
- [61] K.M. Jensen, M.M. Davis, Health care in adults with down syndrome: a longitudinal cohort study, *J. Intellect. Disabil. Res.* 57 (10) (2013) 947–958.
- [62] T.M. O'Lynnner, W.N. Al-Holou, J.J. Gemmete, et al., The effect of age on arteriovenous malformations in children and young adults undergoing magnetic resonance imaging, *Childs Nerv. Syst.* 27 (8) (2011) 1273–1279.
- [63] S.W. Patrick, M.M. Davis, A.B. Sedman, et al., Accuracy of hospital administrative data in reporting central line-associated bloodstream infections in newborns, *Pediatrics* 131 (Suppl 1) (2013) S75–S80.
- [64] K. Singer, P. Subbaiah, R. Hutchinson, et al., Clinical course of sepsis in children with acute leukemia admitted to the pediatric intensive care unit, *Pediatr. Crit. Care Med.* 12 (6) (2011) 649–654.