UNIVERSITY OF CALIFORNIA,
IRVINE

Design, Implementation, and Usability of the Electronic Medical Record Search Engine (EMERSE) Tool

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Informatics

by

Colby Reyes

Thesis Committee:
Professor Kai Zheng, Chair
Associate Professor Yunan Chen
Assistant Professor Daniel Epstein

2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Numbers U24CA204863 and P30CA046592, as well as the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Numbers UL1TR000433 and UL1TR002240. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

I would like to thank Kai Zheng, Yunan Chen, and Daniel Epstein for serving on my thesis committee, as well as David Hanauer and Lisa Ferguson for their guidance in this project.

ABSTRACT OF THE THESIS

Design, Implementation, and Usability of the Electronic Medical Record Search Engine (EMERSE) Tool

By

Colby Reyes

Master of Science in Informatics

University of California, Irvine, 2021

Professor Kai Zheng, Chair

Introduction

Free text forms of clinical documentation stored in Electronic Health Record Systems (EHRs) contain a trove of data for researchers and clinicians alike. However, often this data is not easily accessible for a number of reasons. Many tools have been developed to help users in the task of searching through free text notes in clinical documentation. One such tool is the Electronic MEdical REcord Search Engine (EMERSE), a clinical documentation search tool developed by the University of Michigan. The purpose of this study is to evaluate the usability and user experience of the EMERSE system for a variety of users.

Methods

The study was conducted in 3 phases. In Phase 1, interviews with site administrators investigated factors that facilitate or hinder the implementation and adoption of a system such as EMERSE. Phase 2 employed semi-structured interviews to understand the uses, benefits, and limitations of the system from the perspective of experienced "power users". In Phase 3 system-naive users performed a set of basic workflow tasks, paired with post-activity questions and surveys, to evaluate the intuitiveness and usability of the system.

Results

Users rated the system exceptionally high on usability, user interface satisfaction, and per-

ceived usefulness. Feedback also indicated that improvements could be made in visual contrast, affordances, and scope of notes indexed.

Conclusions

These results indicate that tools such as EMERSE should be highly intuitive, attractive, and moderately customizable. This paper discusses some aspects of what may contribute to a system having these characteristics.

Chapter 1

Introduction

1.1 Clinical Information Extraction Tools

Electronic Health Records (EHRs) have expanded rapidly in use over the last decade. While this has allowed for the generation and storage of vast amounts of clinical data, use of this data for clinical, translational, and epidemiological research, point of care applications, clinical quality assurance, and other purposes has been limited by the form and nature of this data [39, 15]. The natural and expressive nature of free text clinical documentation makes it useful to clinicians in the immediate context of care [39]. However, outside of its original, immediate context and coupled with the sheer volume of data generated in this format on a daily basis, much of this data becomes inaccessible – and therefore unusable – to clinicians, healthcare administrators, and researchers [15, 13, 11, 16]. Conversely, if properly harnessed, broader use of this clinical data could propel the healthcare system towards a true Learning Health System, a self-learning vehicle that through the interrelation of data and information feeds a virtuous cycle of improvement of through cohesive, perpetual interplay between research and application [13, 9, 10]. As a result, various clinical information extraction (CIE) tools have been developed over the years in an attempt to solve this issue

and provide tools that allow users to harness the value of the information contained in clinical documentation for research, clinical care, and other various outcomes. Wang et al [39] present a review of such tools and conclude that the adoption of CIE use is still in its early stages of development, as CIE tools are only used in a small portion of the published papers that use EHR data [39]. They give as reasons for this that access to EHR data for NLP professionals is limited; most clinical NLP relies on the more obsolete practice of rule-based approaches because it can more easily accommodate the knowledge of clinical domain experts; and lack of interoperability that would allow for the training and transfer of CIE tool use across locations and clinical systems. Of note, however, is that the EMERSE system does not appear in their list of CIE tools, despite its use in over 450 peer-reviewed publications and its documented and use [37] in all of the application usage categories identified by Wang et al for NLP-based CIE tools [39]. However, this is likely because their review focused on studies evaluating NLP-based tools and applications and excluded articles on "information retrieval" [39]. However, this exclusion is indicative of the larger trend towards NLP-based tools that may in some respects be more powerful than EMERSE, but generally lack the same measure of usability and user-friendliness [15, 14, 18, 41], or have not yet been implemented in a tool with the usability and utility that EMERSE has demonstrated [37].

While Wang et al reviewed studies evaluating NLP-based CIE tools, toolkits, and frameworks[39], Hill et al have subsequently reviewed the usability and common use of CIE tools, specifically within-EHR search tools, in the clinical setting [15]. While they note that such search tools are generally positively perceived by users, they still suffer from substantial issues in the realms of usability and utility [15]. Furthermore, Hill et al identify medical events, patient treatment/medications, laboratory test results, and allergies as the most critical categories of information that EHR search engines must accommodate to provide sufficient utility for clinical users. In terms of usability, the authors note that these tools impose a high cognitive load on users [15]. Though the authors do not address it, these issues and weaknesses are likely in large part due to their being embedded in the EHR, which has repeatedly

been shown to cause high user stress due to poor usability, including difficult and confusing interfaces, and inability to easily find relevant information [8, 1, 35].

1.2 The EMERSE System

The EMERSE system occupies a liminal space between the types of tools reviewed by Wang et al [39] and Hill et al [15]. This is because, while the EMERSE system does not rely on NLP, it is more akin to the CIE tools discussed by Wang et al [39]. However, while some may categorize it more similarly to the search engine and information retrieval tools discussed by Hill et al, the EMERSE system is not part of the EHR, per se, though it can be and has been integrated into the EHR interface at the University of Michigan. Rather, EMERSE is an information retrieval system, developed at the University of Michigan, designed to facilitate retrieval of information from narrative documents stored in electronic health records. In essence, it is a Google-like search engine with features specifically designed to work with medical text. Due to its design [11], it is suited to clinical and translational research uses like the NLP-based CIE tools reviewed by Wang et al [39], as well as uses in the context of clinical tasks like the tools reviewed by Hill et al [15][1]. Consequently, EMERSE is situated to address a number of issues and weaknesses present in NLP-based CIE systems and within-EHR search tools. A number of studies have introduced and evaluated EMERSE independently, as this study will also do. The development of EMERSE started over a decade ago. To date, it has been used in supporting nearly 1,000 clinical and translational studies and contributed to more than 450 peer-reviewed publications and 82 abstracts, as well as the creation of various other resources (e.g. centralized research registries, repositories, databases, PhD theses, and other initiatives) [37]. This study follows upon an ongoing series of research studies on the EMERSE system, seeking to understand factors that facilitate or detract from the system's

[1]Note that while EMERSE was included in the study by Hill et al, the applications of EMERSE to research were explicitly excluded, and one of the key solutions (collaborative search filtering) to limitations of other current within-EHR search engines is based on a core functionality of EMERSE (collaborative search term bundles)

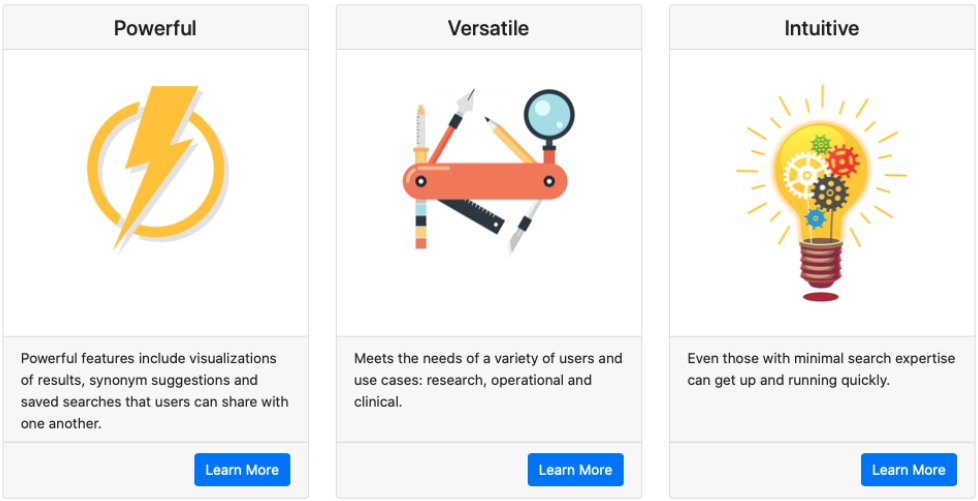


Figure 1.1: The aims of the EMERSE system.

usability and utility in supporting such works.

The first paper, in 2006, introduced the EMERSE system with an overview and validation of it purpose and basic functionality [11]. Over the 15 years since that time, a number of studies have evaluated various aspects of the EMERSE system at different points in the ongoing history of its development [14, 12, 11, 13, 43, 40, 34].

In 2009, Seyfried et al performed one of the first validation tests of the EMERSE system, comparing the speed and accuracy of chart reviews using EMERSE versus performing a manual review of an EHR chart, demonstrating that the EMERSE system produced significant improvements in the efficiency of chart reviews [34]. Most notably, and setting the tone for what has been a theme throughout the history of EMERSE's development, Seyfried et al note that it is not sophisticated algorithms that make EMERSE effective, but rather its focus on a clean and highly intuitive user interface explicitly designed to aid users in accomplishing their information retrieval goals and tasks.

In 2011, Zheng et al published an examination of the use of shared term bundles (groups of search terms shared between users), detailing the extensive degree of sharing amongst a total of 451 users over the course of four years [43, 15]. Sharing of term bundles not

only facilitates the continuity and reproducibility of research [43] and other types of work employing the EMERSE system (e.g. clinical care, quality assurance, billing)[14], but Hill et al also indicate that collaborative term bundles in EMERSE have potential as a means of decreasing cognitive load for clinical users, and build on this functionality to propose collaborative filtering algorithms that suggest synonyms, additional terms, or term bundles based on the user's characteristics [15].

Also in 2011, Yang et al noted that the utility of the EMERSE system may have been limited by the types of queries that the system facilitated the construction of [40]. Furthermore, both of these studies from 2011 conclude that EMERSE users may not all have extensive clinical knowledge, and that the tool must therefore be designed with the needs of such users in mind. That is, such users would not automatically be expected to know all of the synonyms, acronyms, and other types of medical jargon or related terms and concepts to the queries they are responsible for performing [14, 43, 40]. Thus, the system must be flexible enough, powerful enough, and easy to use to facilitate effective queries for users of varying search engine proficiency and varying clinical "literacy" [14, 43, 40]. These insights significantly shaped the development of the system from that point forward, and led to the study undertaken by Hanauer et al in 2017 [14]. In 2015, Hanauer et al published a review of the history of the EMERSE system's development and implementation up to that point [13]. Key takeaways from this study on the usability and acceptance of EMERSE were that, (1) users embraced the shared term bundles feature even more than was anticipated, (2) EMERSE continues to be well-received, perceived as highly valuable for job efficacy and efficiency, and very positively rated by users, as it has been since its introduction [11]. Furthermore, in this article, Hanauer et al identify issues and challenges in meaningfully extracting information from free-text medical records that have guided the way EMERSE has taken shape in the years since. For example, factors like the use of synonyms, acronyms, and abbreviations were noted here [13] and have become a core aspect of EMERSE's utility in its easy-to-use "synonyms" functionality.

In 2017, Hanauer et al published a study piloting a semantically-based query term recommendation functionality for EMERSE. This feature resulted in highly positive feedback from users and also increased perceived usefulness and usability. [14]

In 2019, Zheng et al performed a pilot testing study aimed at understanding how users of the EMERSE system would employ different customization options. They found that over 80% of users did not engage with the customization features, and users who did engage largely customized only the "look and feel" aspects of the system rather than modifying and customizing its functionalities [42].

Finally, in 2020, Hanauer et al performed validation testing [18] to identify how EMERSE has been used to support cancer research [12]. Results of this study indicate not only that the EMERSE system is useful in supporting multiple aspects of cancer research, but also provide some of the first evidence to verify that the system is usable and useful outside of the University of Michigan where it was initially developed and deployed [12].

1.3 Purpose

Figure 1.2: Wordcloud generated from Phase 2 transcripts

As the ONC legislates new norms of interoperability and data access, the EMERSE project is expanding to new sites, and seeks to provide a user-friendly, easy-to-implement clinical information extraction tool to serve the needs of researchers and clinicians in a new era of healthcare interoperability. The purpose of this study is to solicit feedback about EMERSE's usefulness and ease of use, and to identify implementation and usability issues that may not be apparent to the development team.

Chapter 2

Methods

The study consists of three main components, all involving interactions with human subjects. In each component of the study, participants were be research or healthcare professionals, such as clinician scientists, clinical research coordinators, and healthcare administrators, who use EMERSE and/or other research software systems on a daily basis, or oversee such software use and its users. Participation in the study was unlikely introduce any risk to them, and IRB approval was obtained for all three components of the study. Kushniruk et al lay out a timeline detailing when different types of usability studies are appropriate, as seen in Figure 2.1 [18]. Yen and Bakken also lay out a set of principles to guide usability studies for different phases in a system's development lifecycle [41]. These principles were drawn upon in the consideration and construction of the protocols for each phase of this study. Per the framework laid out by Kushniruk et al [18], the Phase 1 interviews can be considered both "Exploratory Tests" and "Assessment Tests", while the Phase 2 interviews fall squarely into the category of "Validation Tests", and the Phase 3 interviews can be classified as "Assessment Tests".

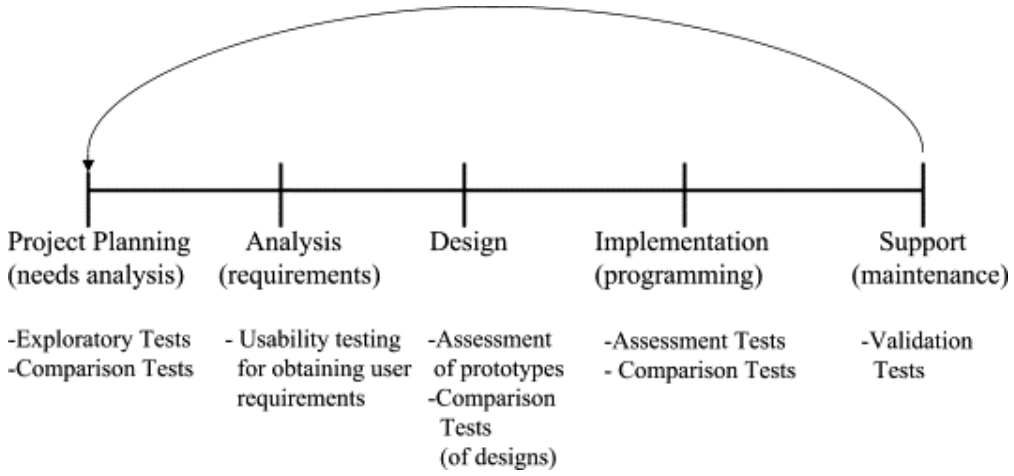


Figure 2.1: Appropriate Usability Studies for Development Phases[18]

2.1 Participants & Recruitment

Study participants consisted of a total of 46 individuals from 7 different sites, including the University of Michigan and 6 additional external project sites. A total of 10 individuals composed the 6 teams from the external project sites for Phase 1, while 20 participants were interviewed individually from the University of Michigan for Phase 2 and 16 participants across 4 of the 6 external sites were interviewed individually for Phase 3.

2.1.1 Phase 1 Recruitment

Participants for the first phase of the study consisted of site leaders and administrators at locations where the administration was either considering, or in the process of, implementing EMERSE as a tool for their staff. One or more individuals, constituting the project teams from each of six different academic medical centers (each in a different state in the United States) involved in the process of implementing EMERSE for their location, were contacted and recruited to participate in a semi-structured interview.

2.1.2 Phase 2 Recruitment

For Phase 2 of the study, 20 veteran EMERSE users at the University of Michigan were identified and recruited by the PI of the larger NCI-funded project, Dr. David Hanauer. Users were identified by system log analysis, pulling the identities of the top users from the EMERSE system logs and working down this list in descending order of total EMERSE usage to recruit users until 20 participants had been scheduled. Furthermore, these participants were engaged users ("power-users") who have contributed to the development, testing, and evaluation of the EMERSE system in the past. One acknowledgement of the study's limitations is relevant here. While this method of recruitment allowed for the recruitment of participants with the greatest amount of experience and familiarity with the EMERSE system, its interface, and its functionalities, this may have increased a potential positivity bias since users who dislike a system are less likely to spend large amounts of time using it.

2.1.3 Phase 3 Recruitment

For the third and final phase of the study, following EMERSE's initial installation and functional testing, five to six prospective users at each collaborating site were identified and recruited by the site PIs to participate in additional usability studies to reveal potential issues that may arise from the sites' local contexts. These participants were clinician scientists, clinical research coordinators, or healthcare administrators, who frequently perform medical chart review tasks that can be facilitated by the use of the EMERSE system.

2.2 Data Collection

In all three phases, data collection took place in the context of interviews conducted via Zoom video-conferencing, with an additional administration of QUIS and TAM surveys for Phase 3 of the study.

2.2.1 Phase 1: Semi-Structured Interviews

Between August and October 2020, we conducted leadership interviews with the project team at each participating site. In most of these interviews, the site PI, in addition to multiple programming and coordinating staff, were present. All interviews took place on Zoom; each lasted approximately one hour.

Proceedings of the interviews followed a semi-structured interview protocol that was developed for this study, found in Appendix A. The protocol contains a total of 12 questions, covering a variety of topics including organizational context, characteristics of anticipated users, progress of implementation, obstacles encountered, and satisfaction with the technical support provided by the University of Michigan team.

2.2.2 Phase 2: Semi-Structured Interviews

Between March and June of 2021, interviews were conducted with 20 veteran EMERSE users at the University of Michigan who have been using the system as a routine part of their job. These interviews were semi-structured, and included questions informed by technology acceptance theories. The purpose of the questions was to solicit feedback about the system's usefulness and ease of use, and to identify usability issues that may not be apparent to the development team. The interview protocol can be found in Appendix B. Each interview lasted between 30 and 45 minutes; all were conducted via Zoom. Upon the participant's consent, the interviews were tape-recorded, and transcribed to facilitate subsequent qualitative analyses. No identifying information was collected during the interviews. Any potential identifying information accidentally disclosed was removed from the transcript. The audio tapes were destroyed after transcription.

2.2.3 Phase 3: Assessment & Validation Tests

User Testing & Contextual Interviews

Between September and November 2021, we conducted user testing and contextual interviews with 16 different participants from 4 different sites. These users were asked to review the user documentation, and then followed a structured test script to perform a set of simulated search tasks. The interviewer collected and catalogued qualitative data in the form of observations[4] during the users' performance of the prescribed tasks. Participants were then asked to report their perceptions and experience with the system using the Questionnaire for User Interface Satisfaction (QUIS), a validated and widely used usability survey instrument (Appendix C, Table C.3.1). They were also asked to fill out a validated questionnaire instrument based on the technology acceptance model (Appendix C, Table C.3.2), which assesses key determinants of technology acceptance behavior among prospective users. Each usability testing session, including the time needed to respond to the questionnaires, lasted between 30 to 50 minutes. No pieces of identifying information were collected. The full interview and demonstration protocol can be found in Appendix C in Tables C.2.2, C.2.3, and C.2.4.

Survey Administration

Survey data was collected by means of two recognized and externally validated surveys: the Questionnaire for User Interface Satisfaction (QUIS)[3] and Technology Acceptance Module (TAM)[22, 19, 38] surveys.

2.3 Data Analysis

2.3.1 Data Preparation

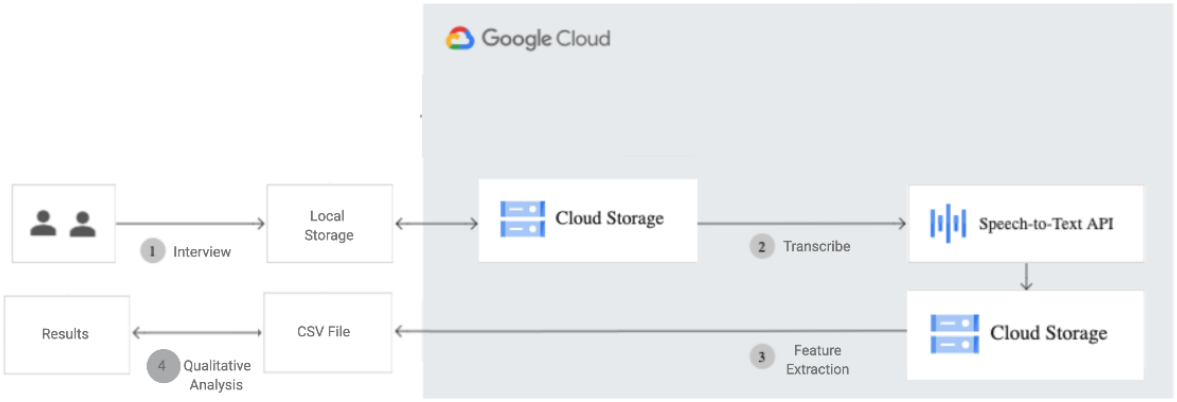


Figure 2.2: Transcription and Qualitative Analysis Workflow

In order to perform qualitative analysis, interview sessions were first transcribed. Transcription was done using the Automatic Speech Recognition (ASR) module in the Google Cloud Platform (GCP). The transcription process followed the workflow as seen in Figure 2.2. Information on the code and repository can be seen in Appendix D.1. Once transcription was completed, interviews from all three phases were coded using the constant comparative method of qualitative coding. Quantitative analysis was performed by means of descriptive summary statistics on both the survey data and the catalogued events and outcomes of the steps of the demonstration scenarios. These analyses are further detailed below.

2.3.2 Qualitative Analysis

As Creswell notes[4, 5], "data collection and data analysis must be a simultaneous process in qualitative research"[4]. Consequently, initial phases of analysis were performed *in situ* as interviews were conducted. This is particularly relevant to Phase 3 of the study, as anomalies and patterns were identified by observation as participants executed the prescribed tasks of the user evaluation. These observations were then further refined and categorized into the

event types laid out in Table 2.1. These outcomes were then funnelled into the quantitative analysis further described below.

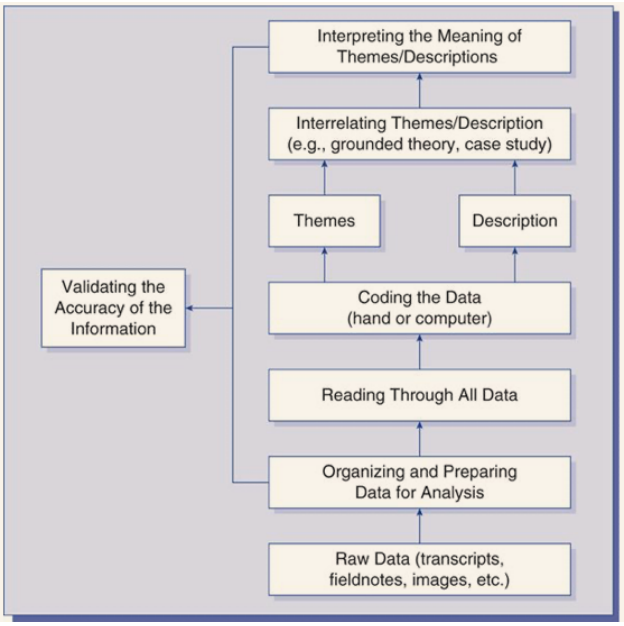
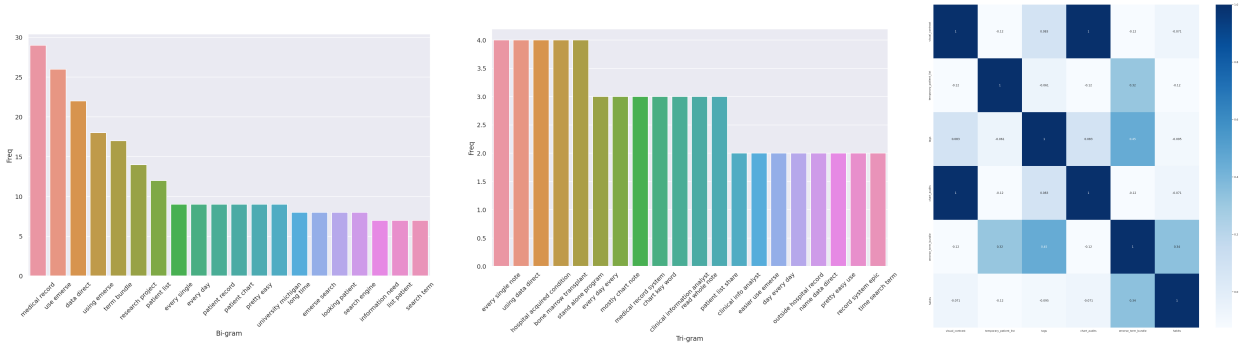


Figure 2.3: Diagram of Qualitative Coding Process

Transcripts from all three phases of the study were analyzed qualitatively following the general procedure in Figure 2.3 posed by Creswell [4] and aligned with the Constant Comparison Method. Following in this process outline, and drawing on tools and techniques from phenomenology (coding significant statements)[4] and grounded theory (open coding and selective coding)[4], each interview transcript with its corresponding field notes was coded by identifying and chronicling significant statements, the interviewer's observation of participant's sentiment toward the software, emerging categories of statements and observations, and the interconnection of codes. Codes were then iteratively refined and re-compared to the data as ideas, patterns, and themes emerged within study phases and across study phases. These ideas, patterns, and themes were chronicled throughout the process and their interpretations are presented in the following Results and Discussion sections.



(a) Top bi-grams for Phase 2 interviews

(b) Top tri-grams for Phase 2 interviews

(c) Correlations of text features from semi-automated coding

Figure 2.4: Example N-grams and Correlations from Semi-Automated Feature Extraction

Additionally, drawing on NLP-based semi-automated qualitative coding techniques used in the analysis of other types of applications and software[17, 20, 21, 23, 6], the qualitative analysis process was aided by the performance of feature extraction from transcripts using NLP and manually grouping overlapping extracted features.[1] For example, NLP-based extraction of the most commonly occurring bi-grams and tri-grams (see Figures 2.4a and 2.4b, respectively) was performed and their respective collocations were analyzed. The number of occurrences of each bi-gram or tri-gram, or manual grouping of similar n-grams, in each transcript was used to generate an array of features and feature counts that allowed for the production of a correlation heat-map (see, for example Figure 2.4c). A full depiction of the different heat-maps as well as top bi-grams and tri-grams used to inform qualitative analysis on all three phases of the study can be seen in Appendix F.3. High-frequency n-grams and the mappings of n-gram co-occurrences were used to aid in the production, examination, refinement, and verification of qualitative codes in the constant comparison method process. In this process, the codes were summarized into larger recurring themes.

[1]Notebook code can be found at `https://github.com/colby-reyes/EMERSE/blob/main/EMERSE_FeatureExtraction.ipynb`

2.3.3 Quantitative Analysis

User Testing Outcomes Statistics

For each step within the three scenarios of the user testing protocol (Appendix C), 6 different types of events or outcomes were catalogued as participants completed the prescribed tasks. These event types consisted of one *"Fail"* type, one *"Help Request"* type, three *"Confusion"* types (`Searching`, `Delay`, and `Initial Mistake`), and a *"No Issues"* event type. Full descriptions of each event type can be seen in Table 2.1. Of these event types, *"Failure"*, *"Help Request"*, and *"Confusion"* events were not mutually exclusive. That is, within a single step, a participant may have exhibited `Searching` and had a *"Help Request"*. Alternatively, for example, if a participant made a mistake in the execution of a task that they did not recover from by themselves, and then asked for help to complete the task correctly, this would have been catalogued in both the *"Fail"* event type and *"Help Request"* event type. Only the "No Issue" event type is exclusive, and indicates that the step was completed without the occurrence of any other event types. Descriptive statistics were used to analyze the occurrence of different types and combinations of catalogued events.

Event Code	Type	Description
Error/ Failure (F)	*Fail*	Participant "failed" the step (completed it incorrectly, or skipped it)
Help (H)	*Help Request*	Participant asked for help completing the step or help understanding what a step accomplished
Initial Mistake (I)	*Confusion*	Participant initially made a mistake (e.g. wrong click) and then corrected it themselves
Delay (D)	*Confusion*	Participant exhibited a notable delay between actions
Searching (S)	*Confusion*	Participant exhibited behaviors (e.g. mouse movements) indicative of searching for interface items
No Issue	*No Issue*	Participant completed step with none of the above events

Table 2.1: Event types

Survey Data

Data from each survey was analyzed using descriptive statistics.

Chapter 3

Results

3.1 Qualitative Analysis

3.1.1 Phase 1

Overall, the feedback from all sites about the project was very positive. All sites were excited about the prospect of using EMERSE to meet the local needs for retrieving information from free-text medical records. All spoke highly of the ease of deploying the software and the responsiveness of the University of Michigan supporting team. Most improvement suggestions were made by early adopting sites and focused on the inadequacy of software documentation and deployment instructions. These issues have been addressed since, as the sites that started implementation later in the project did not report the same observations. Other improvement suggestions included accommodation of additional database management systems besides Oracle, and a few user interface enhancements to facilitate local workflow.

Theme: Barriers to Implementation

When asked about key barriers to implementation, most project teams reported that obtaining organizational approval was challenging and was often the main reason accounting for implementation delays. Except for Columbia University, none of the other sites had prior experience in using software systems for extracting information from free-text clinical data. All sites commented that the process of installing and configuring EMERSE was straightforward. The only technical barrier was to import clinical notes from the electronic health records system and format the notes properly for consumption in EMERSE. Overall, we did not observe any significant concerns regarding the project despite anticipated delays at some of the sites.

3.1.2 Phase 2

Qualitative analysis of Phase 2 interviews surfaced key themes relating to (1) the desire for the system to index more data sources so that EMERSE could be used for more tasks and/or without the need for any other software interactions, and (2) the need for promotion of the system to expand the system's user base.

Theme: Access and Integration

A major theme that emerged from interviews with experienced users was the desire to be able to use EMERSE for even more tasks and on more sources of data. Users indicated that reliance on another tool (e.g. the EHR, RedCap, DataDirect) to get to data that EMERSE does not index slowed their work and hindered usability. Integration, in one form or another, with more sources of data was positioned as a means to make the tool even more powerful and useful to them in their work. The desired forms of integration fell into three general categories: discrete data (labs, vitals, etc), scanned documents, and external site records.

Discrete Data

A large proportion of users interviewed in this phase noted that some form of integration with the EHR's database of discrete variables would make EMERSE even more powerful and increase its utility for them. This database of discrete variables includes lab results, vitals, and other tabular data stored in the EHR that EMERSE currently does not index because it is limited to free-text data. While EMERSE is accessible through the EHR at the University of Michigan (where these participants were recruited from) by the integration of a button that takes users from the EHR to EMERSE, participants noted that reliance on another tool to get discrete data often slowed their workflow and presented the greatest limitation to the usability and utility of EMERSE in their work. Furthermore, though some lab documents are included in the "Scanned/PDFs" portion of the results, they are not sufficient for the purpose desired by participants, due to both their quality (which is substantially lower than the rest of the results seen in EMERSE) and their lack of comprehensiveness (i.e. they only reflect lab reports that had a document to be scanned in and were chosen to be scanned in and therefore they do not represent the totality of each patient's discrete data).

Reports and Scanned Documents

Users also repeatedly noted the desire for EMERSE to index and present results from more document types that are currently not supported. Among these desired document types are reports such as the text attached to or embedded in CT, MRI, EKG, and other reports. While these types of notes are similar to those seen in the Radiology Notes category of EMERSE results, these requests indicate that the current scope of Radiology Notes indexed is too limited. In addition, users also desired expanded access for scanned-in documents. As noted above, EMERSE does contain a category for scanned documents and PDFs, but this recurring theme from participant feedback indicates current limitations in regards to one or both of (1) the data entry and health information management processes by which these documents are added to EMERSE's database, and (2) the capacity of Optical Character Recognition (OCR) tools to extract and transform data from scanned PDF images into searchable text.

External Site Records

Finally, within this theme, participants repeatedly expressed desire for the expansion of the network of patient records from which EMERSE was able to draw. That is, they desired integration with and expansion to external data sources. University of Michigan users noted that this is partially supported by "CareWeb" (this is also supported by the "Network" feature in the available lists of patients in EMERSE), but even with this integration and the size of the University of Michigan health system, users still found themselves limited by the inability to access records from external sites and healthcare systems that were not part of CareWeb. Participants also suggested that still more external sources of data would improve the system's utility. Notably, users desired an integration with data from RedCap. It is important to note that this request for access to documents and records from non-network clinics reflects the current state of healthcare interoperability. This will be further examined in the discussion below.

Theme: Promotion of the System

Users unanimously responded that they would recommend the system to others, and repeatedly noted that while they found the system exceptionally useful and easy to use, the distribution of its use and awareness of its existence – even within the University of Michigan where it was developed – was not proportionate to its utility. That is, for a tool they found so useful, they believed that far too few people both within and outside of their network knew about the tool's existence. Some users even went so far as to suggest that if large national healthcare or insurance companies knew about EMERSE, they would certainly implement it as their main tool for data abstraction.

3.1.3 Phase 3

Qualitative analysis of observations and participant questions during system testing, transcripts of post-testing interviews, and participant responses in survey comment fields revealed

four major themes regarding the participants' perception of and experience with the system. These themes were, (1) the expectation of (and reasons for) a brief, but necessary adaptation period to become proficient with use of the system, (2) the utility the participants expected the system would provide in their respective clinical or research work roles, (3) notes about the use and role of colors in the system, and (4) clarity and resetting of patient lists.

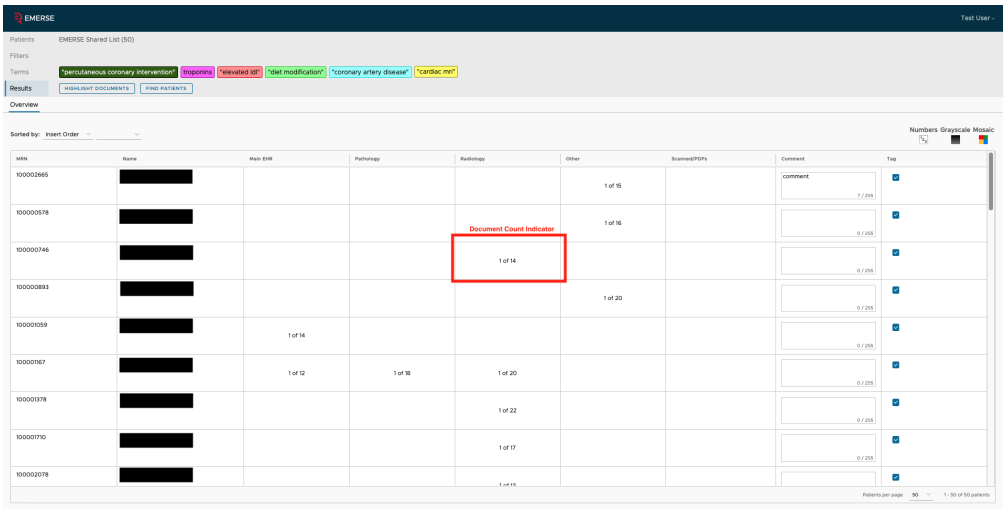


Figure 3.1: The *Highlight Documents* view

Theme: Adaptation Period

A recurring theme in testing with system-naive users was that they expected to find the system even easier to use after a brief adaptation period. Many noted that even simply retaining the instructions for the basic scenarios used in this phase for one or two more uses would be sufficient to become completely familiarized with the interface and the functionality of its components. Key hindrances to immediate intuitiveness (i.e. the reasons given that a brief "adaptation period" would likely be necessary to become satisfactorily proficient with the interface) were related mainly to navigation and the apparentness of affordances of system elements. For example, the ability to click a cell with a document count indicator in the table as seen in Figure 3.1 and be taken to an overview of that category of documents was an affordance that was not immediately apparent to a substantial number of users. A similar

issue was noted with the `Patients`, `Filters`, `Terms`, and `Results` navigation elements, as seen in the top left of Figure 3.1.

Theme: Expected Utility

During interviews, participants repeatedly expressed their perception that the EMERSE system would be highly useful to them. They also repeatedly expressed their desire to start using the system in their real work, with statements such as:

> "Right now we have to request from a department within the system to run it for us. Okay, if we could do this ourselves, it would be amazing."

> "How soon can we start using it?"

Theme: Use of Colors

Users in both phases 2 and 3 made notes pertaining to the accessibility of the user interface, mainly in regards to colors and visual contrast. The majority of these notes were positive, while some presented useful critiques to help improve system accessibility. Positive feedback on the use of colors in the EMERSE interface fell into the categories of interface aesthetics and color-coding. A number of users noted that the color scheme and flat design of the EMERSE interface made the system very visually pleasing and facilitated finding information and understanding the organization of information. Furthermore, color-coding of search terms according to the Boolean logical relation of the terms (i.e. terms joined by the "`OR`" Boolean logic – or synonyms – are displayed with the same color while terms separated by the "`AND`" Boolean logic are distinguished by different colors) serves to decrease cognitive load on users in the performance of the search task. Users noted that such color-coding increased their understanding of the system's grouping of information and improved its visual distinction and recognizability in reviewing the interface for information and the occurrence of different terms in documents during queries. Critiques of the use of color fell into the category of visual

contrast, with a subset of these expressing the need for ADA-compliant colors in particular. Participants expressed that lower visual contrast made it more difficult to understand the function of different aspects of the interface. Of note is that this contributed especially to the non-obvious nature of the affordances of the `Patients`, `Filters`, `Terms`, and `Results` navigation elements. Users also suggested that increasing the visual contrast of the `Find Patients` and `Highlight Documents` buttons when they are available to click would increase the usability of these core features of the system by making their affordances more obvious. Finally, the use of ADA-compliant colors and color contrast schemes was an issue raised by numerous users in phases 2 and 3. The implications of these notes are further discussed below in relation to the principles of Accessible Design.

Theme: Patient Lists

The tasks related to patient lists, while mainly related to navigation elements in the system interface, emerged as a separate theme from the navigation elements for two reasons: (1) these are issues that veteran users also noted (and thus did not fit into the category of navigation elements users believed they "just need to get used to"), and (2) understanding of patient lists is central to the EMERSE workflow. While users perceived the importance of patient lists, they raised concerns related to awareness of the current patient list, changing between patient lists, and resetting search parameters for a new search across all available patients.

3.2 Quantitative Analysis

Results from quantitative analysis indicate that new users of the EMERSE system are able to complete basic but critical workflow tasks in the system with a high rate of success (low error rate and low rates of confusion or needing help), are highly satisfied with the interface, and have highly positive perceptions of its expected utility in their work.

3.2.1 User Testing Outcomes

Analysis of the number of times users failed a step, asked for help, or exhibited behaviors indicating confusion with the interface (i.e. making a mistake and subsequently correcting it, mouse movement indicating searching for an interface feature, and/or a notable delay between steps or actions) can be seen in Tables 3.1, 3.2, and 3.3[1]

Noteworthy Steps

The most "troublesome" step overall was Scenario 1, step 7, with 31% of users exhibiting a delay or searching behavior and only 56.25% of users completing the step with no issues (see F.2). Connecting this outcome to the qualitative analysis reveals that this searching was due to both difficulty finding the location of the *Highlight Documents* button and confusion over the "clickability" and/or meaning of the document count indicator (i.e. the `"# of #"`, as seen in 3.1)in the columns of the *highlight documents* view. The second-most "troublesome" step was Scenario 3, step 2. In this case, only 68.75% of users completed the step without issue, with 25% of users exhibiting delays and 18.75% exhibiting searching behaviors (see F.2). Cross-referencing with qualitative analysis, this is an example of the "navigation" theme feedback, the issue in this case being finding the *Saved Terms* tab within the *Terms* view. The step with the highest error rate was Scenario 2, Step 1, with 2 (12.5%) of the participants having an error on this step. This step involves the removal of a search term as the first step to start a new query after completing a prior query. In connection with the themes that emerged from qualitative analysis of Phase 3 interview responses, the higher error rate on this step indicates that the affordance of the pencil icon in the corner of each search term is not immediately apparent as the means by which to remove that specific term. In connection with this, specific feedback from users indicated that, because the pencil icon also allows for editing of the term without removal of it, the addition of a button on each term such as an "x" or a "trash can" icon would be useful to separate the action of removing the

[1]Analysis of outcomes by individual task for each scenario can be seen in Table F.2

Scenario	Errors	Help Requests	Search/Delay	Corrected Mistakes
1	1.56%	3.13%	11.72%	3.13%
2	1.56%	2.60%	11.46%	3.13%
3	2.08%	3.13%	18.75%	5.21%
Overall	**1.74%**	**2.95%**	**13.98%**	**3.82%**

Table 3.1: Summary of "negative" event occurrences in user testing scenarios.

term from the action of editing it, thus making the option for the user to remove individual terms more obvious.

The steps with the highest help request rate were Scenario 1, step 7 and Scenario 2, Step 4, with 2 (12.5%) of the participants requesting help on each of these steps. Scenario 1, Step 7 represents the first instance of the use of the `Highlight Documents` button, as well as the instruction to click on a specific cell containing a document count indicator (see Figure 3.1 for reference). Based on qualitative analysis of interview responses, both of these interface elements were a source of confusion for users due to lack of visual distinction for the `Highlight Documents` button and the functionality of the cells in the `Highlight Documents` table as links not being immediately apparent. Scenario 2, Step 4 represents the first time users are asked to explicitly change the `Patient List`. The requests for help were due to both an inability to find where options to explicitly change the patient list were located and confusion over why this was a necessary step to proceed with the new query executed in Scenario 2. Qualitative analysis above and further discussion below elucidate why this is the case.

Of note is that 77% of *all steps* were completed by *all users* with *no errors*, 53% of *all steps* were completed by *all users* with *no help requests*, and 31% of *all steps* were completed by *all users* with *no issues*. Across all steps of the three scenarios, as summarized in Tables 3.1 and 3.2 the average error rate was 1.74%, the average help request rate was 3.8%, and the average rate of completion with no issues was 88.3%.

Scenario	*No **Errors***	*No **Help** or **Errors***	***No Issues***
1	98.44%	95.31%	*89.84%*
2	98.44%	95.83%	*89.58%*
3	97.92%	94.79%	*85.42%*
Overall	**98.26%**	**95.31%**	***88.28%***

Table 3.2: Summary of "*Fail*" and "*Help Request*" event **non-occurrences**.

Scenario 1	**Scenario 2**	**Scenario 3**	***Overall***
89.84%	89.58%	85.42%	***88.28%***

Table 3.3: Percent Step Completion with *No Issues* in user testing scenarios.

3.2.2 Surveys

QUIS

QUIS survey results indicate very high ratings of usability and satisfaction with the EMERSE system by new users after only one session of use. On the scale of 0 through 9 on the QUIS survey, users on average rated their overall impression of the software at 7.47 ± 1.33. The lowest-performing question was the ranking of the sequence of screens on the scale from "confusing" (O) to "very clear" (9), with an average rating of 6.31 ± 1.89. Meanwhile, the highest-performing questions were both in relation to system capabilities, asking about system speed (rated at 8.8 ± 0.40) and quality of search results (rated at 8.67 ± 0.72). This is especially of note since these are two of the core features of the EMERSE system. Also of note is that the highest-performing aspects of the EMERSE system based on the results of the QUIS survey were the "Usability & UI" and "System Capabilities", rated at 7.94 ± 1.21 and 8.19 ± 0.96, respectively. Average scores of each question and each question group can be seen in Figure 3.2.

TAM

TAM survey results indicate high acceptance of the EMERSE system. To facilitate analysis, Likert Scale ratings were assigned the numerical values of -2 for `Strongly Disagree`, -1 for `Disagree`, 0 for `Neutral`, 1 for `Agree`, and 2 for `Strongly Agree`. The average score

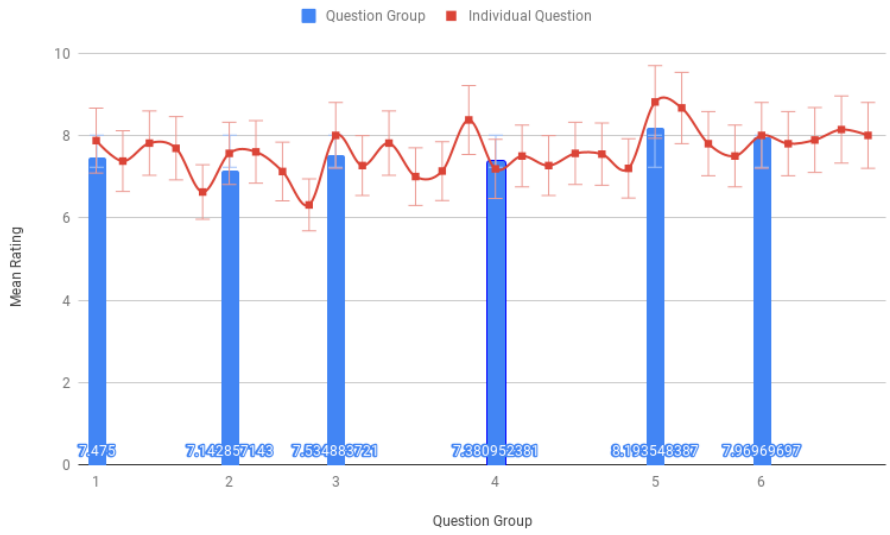


Figure 3.2: QUIS Responses by Question Group

across all 10 questions was 1.47 ± 0.13. Average scores for each question can be seen in Table 3.2.2. Also as seen in Table 3.2.2, the highest performing questions were the final two, asking whether users would find it easy to become skilled at using the system and whether they would find the system easy to use. These questions were rated at 1.69 ± 0.48 and 1.63 ± 0.50, respectively. The lowest rated question was in regards to whether users believed the system would improve their job performance, but was still rated at 1.31 ± 0.79.

Table 3.4: TAM survey ratings by question

Q#	**Question**	**Average Rating**
1	Using this system in my job would enable me to accomplish tasks more quickly.	1.56
2	Using this system would improve my job performance.	1.31
3	Using this system would enhance my effectiveness on the job.	1.38
4	Using this system would make it easier to do my job.	1.50

Continued on next page

Table 3.4 – *Continued from previous page*

Q#	Question	Average Rating
5	I would find this system useful in my job.	1.40
6	Learning to use this system would be easy for me.	1.44
7	I would find it easy to get this system to do what I want it to do.	1.31
8	My interaction with this system would be clear and understandable.	1.50
9	It would be easy for me to become skillful at using this system.	1.69
10	I would find this system easy to use.	1.63
	Overall	***1.47***

A summary of the ratings across all questions of the TAM survey can be seen in Figures 3.3 and 3.4. Figure 3.3 displays the percent of answers across all questions that fell into each Likert scale category, while Figure 3.4 displays the average rate at which participants chose each Likert scale option across the 10 questions of the survey. For all 10 questions in the TAM survey, the option `Strongly Agree` indicates the highest level of acceptance of the software.

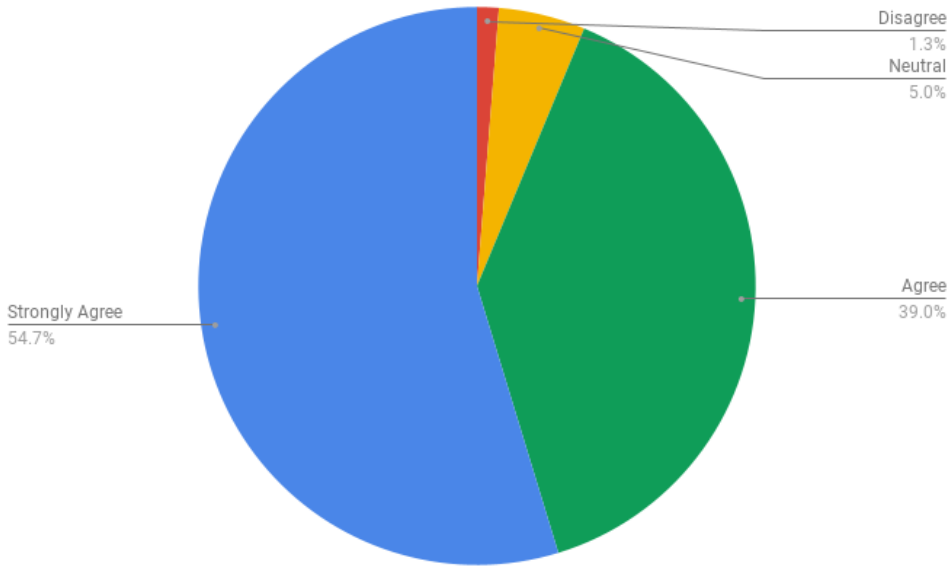


Figure 3.3: Summary Measures of Likert Scale responses for TAM Survey

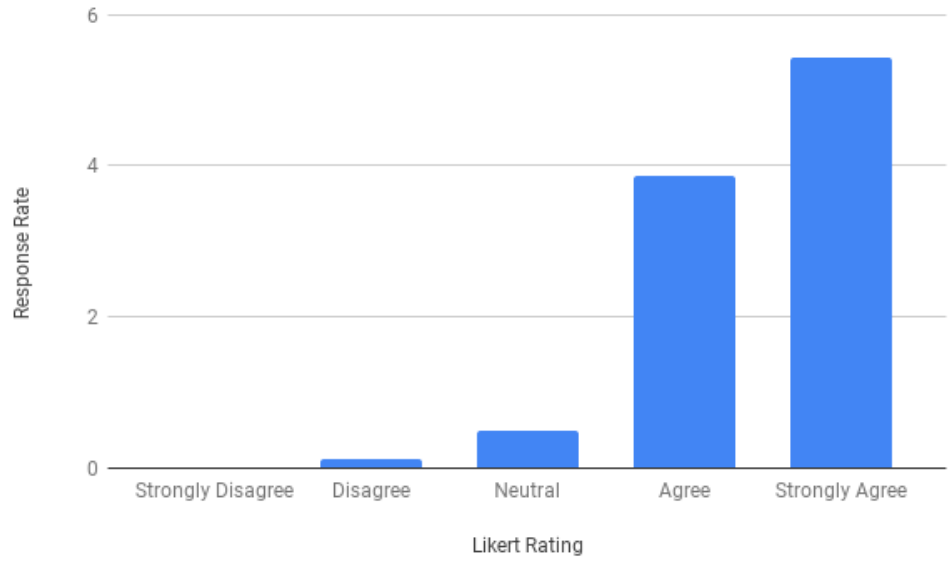


Figure 3.4: Summary Measures of Likert Scale responses for TAM Survey

Chapter 4

Discussion & Conclusions

4.1 Discussion

4.1.1 Utility

While Hill et al assign the task of overcoming limitations of current EHR-based search engines to NLP and AI techniques, EMERSE is able to overcome a number of current issues related to search engine result quality through its query methods [14] and effective, highly intuitive design without relying on NLP. For example, Hill et al posit that it will require NLP and AI implementations to solve the issue of search term permutations such as misspellings, acronyms, and synonyms so as not to miss results related to what the user may have intended or expected to include. However, this approach suffers from the inability to restrict the search back to a very literal search for the exact term input by the user, while EMERSE accommodates for these permutations through extensive and easy-to-use synonym lists and spell-check suggestions. This approach taken by EMERSE provides the flexibility to expand the search to accommodate exactly what the user's mental model of a search term may include without losing the assurance of result accuracy that intuitively arises from

literal string matching. It is worth noting here, though, that the cognitive burden on users in cases where they may desire to employ term permutations may be further decreased by implementing some form of term or term bundle suggestion mechanism, whether it be by a collaborative filtering algorithm as posed by Hill et al [15] or a semantically-based query recommendation as explored by Hanauer et al [14].

There is one issue that still persists and has not been solved through the design of EMERSE and may require an NLP-based solution. This issue was noted by participants and relates to Hill et al's discussion of search engine results [15]. This is the issue of negations [15], in which a search engine like EMERSE with highly literal results may return a "positive" hit for a term when in fact the text in question contains context that would negate that hit as a true positive. For example, searching for patients whose charts contain the term *"Coronary Artery Disease"* may return as "false positives" patients whose charts include the literal string match, but actually read as *"no family history of Coronary Artery Disease"* Consequently, a solution to this issue may still require the use or integration of an NLP-based functionality.

4.1.2 Design & Usability

Affordances & Usability Heuristics

Before continuing, it is worthwhile to summarize some basic principles in the realm of design and usability, as these principles have informed not only the design of the EMERSE system, but also informed the construction of the evaluation tools and methods used, and will be used as criteria to inform the analysis, evaluation, and discussion of user feedback. Two authors have been formative in this area, Don Norman and Jakob Nielsen, and their work will be briefly summarized. Don Norman's idea of affordances is a key principle in software engineering and interface design, and is also central to the the understanding of usability evaluations. Norman defines an "affordance" in the context of software interfaces as what

an interface item visually signals to the user that it is able to "do" [32]. Norman also popularized the idea that more attractive system designs stimulate positive emotions that then increase users' perception of the usability of an artifact [32, 31, 30, 24]. These principles are built upon by Jakob Nielsen, and he has outlined 10 design heuristics for system usability [29, 27, 28, 26][1] that are often useful as general criteria for evaluation of system functionality, and serve as touchstones for evaluating user feedback. In this paper, themes and outcomes from the three different phases of user evaluation will be discussed and evaluated in light of these criteria. Nielsen's 10 heuristics are as follows[26]:

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and Standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation

Results of qualitative analysis are discussed in light of these principles.

[1]Descriptions of each heuristic can be found in Table F.1

Accessible Design

User feedback related to colors, visual contrast, and visual distinction is largely aligned with the increasing awareness and concern for accessible and inclusive design[44, 7, 25, 33, 2]. As other websites and companies have sought to comply with ADA guidelines for internet accessibility, awareness of the benefits of accessible and inclusive design has trickled down to users who may not necessarily *"require"* accessible design. This is consistent with a key idea of accessible design: designing for inclusion at baseline will benefit *all* users[44, 7, 25, 33, 2]. As a positive example, a feature in EMERSE that hearkens to this principle is the color-coding of terms. This features has shown highly positive reception among users not only because it improves visual contrast and clarity of information presented on the screen, but also serves to clarify the logic of Boolean search terms that users less familiar with advanced search engine functionality [14] may not intuitively understand. Furthermore, this feature indicates the potential of color-coding and other similar uses of system colors in clinical information software to comply with Nielsen's 6^{th} heuristic, which stresses the value of recognition over recall. In addition to the benefits of color-coding, the interface stacks up well against Nielsen's 8^{th} heuristic. As noted in the above section, positive feedback on interface colors and flat, minimalist design reflect the validity of Neilsen's 8^{th} heuristic in the evaluation of clinical software tools.

Patient Lists and the Importance of a `RESET` Button

The issues and confusions noted by users surrounding patient lists can be best understood and evaluated in light of Nielsen's 1^{st}, 3^{rd}, and 4^{th} heuristics [26]. It will be most useful here to start with examining Nielsen's fourth heuristic: **Consistency and Standards**. Here it is noted that users are highly influenced by the norms of designs and functionalities seen in the numerous other systems they interact with on a daily basis in their personal and professional lives and "should not have to wonder whether different ... actions mean the same thing". While it is acknowledged that "EHR search is substantially different from Web search"[40],

the persistence of patient lists across the changing of search terms nevertheless may not be intuitive to users because of the expectations set by other common search engines, perhaps all the more so because EMERSE is often promoted as being a "Google-like" search engine. That is, traditional search engines typically default to search across all available sources unless otherwise specified, whereas once a patient list is added or selected in EMERSE, it persists until it is explicitly removed. This brings us to the issues related to Nielsen's first heuristic: **Visibility of System Status**. Although in regards to patient lists EMERSE technically does meet the criteria set forth in this heuristic by keeping a display of the name and patient count of the current patient list at the top of screen (see Figure 3.1 to find where this is displayed), it may be that this display does not *functionally* fulfill the criteria of this heuristic. That is, in order to supersede user assumptions, this display may need to be larger and "louder". This brings us finally to the third heuristic: **User control and Freedom**. In this heuristic, Nielsen highlights the importance of *Undo*, *Redo*, and *"Emergency Exit"* capabilities of the system. While EMERSE does this quite well in some other respects, participants often requested the addition of a "reset" button or buttons, either to completely re-start their search, or to reset patient lists, filters, and/or search terms.

Customization

While the theme of customization was not substantially explored in qualitative analysis, it does nonetheless pose an interesting point of conversation. While the system's simple, intuitive, and minimalist UI, in alignment with Nielsen's 8^{th} heuristic, likely minimized user desire for customization, customization requests are worth noting in light of Zheng et al's paper on the utilization rates of system personalization in the very same context of the EMERSE system [42]. While some customization requests were unique (e.g. feature request from a Phase 3 participant for custom sorting options in the *Highlight Documents* view) and would likely fall into the the greater than 80% non-utilization rate, the most common customization-type requests in interviews were actually requests for features that already

existed (e.g. choosing the colors with which terms and term groups are color-coded). The recurrence of participant statements and requests like this feeds into the theme of non-obvious affordances that we will now discuss more in detail.

Navigation and Non-Obvious Affordances

In this study, hindrances to immediate intuitiveness of the EMERSE system were related mainly to navigation and the apparentness of affordances of system elements. One of the major themes that emerged from the data and results was the idea of non-obvious affordances of interface elements. While the hidden nature of a number of shortcuts and customization options plays well with Nielsen's 7^{th} and 8^{th} heuristics of "Flexibility and efficiency of use" and "Minimalist design" that allow for shortcuts, customizations, and optimizations for advanced users, often these affordances were too well hidden (and hidden in plain sight) in the EMERSE system from users who desired to find them. Furthermore, the feedback related to the visual contrast of the `Find Patients` and `Highlight Documents` buttons highlights the importance of Nielsen's 1^{st} exemplifies in software interfaces in general and in clinical software tools in particular.

4.1.3 Interoperability

The desire repeatedly expressed by participants for expanded scope of document access is also highly reflective of the current state of healthcare interoperability in the United States, as information blocking practices and the segmentation of EHR systems and clinical databases is highly prohibitive of the free but secure flow of medical record information between the EHRs of different healthcare systems or clinic locations. However, tools like EMERSE with exceptional usability and user-friendliness that promote their utility for users, will become all the more important and necessary as interoperability legislation expands the scope of data upon which clinical databases (and consequently search engines and other CIE tools) can draw, since the increased capacity for data does not inherently bring with it the increased

availability of useful information from that data [16, 14, 13, 11].

4.1.4 Limitations

One limitation of this study, as seen in the large standard deviations of the average survey scores and quantification of user testing outcomes, is the low power of Phase 3 of the study due the small number of participants. Another important limitation of note for this study is that recruitment of participants was not completely random. Future studies could perform more random recruitment of participants for the types of testing done in Phases 2 and 3 of this study to decrease potential "positive user" bias that this study may have been subject to. However, this was a tradeoff made intentionally in this study: while high-frequency users (phase 2) and semi-volunteers (phase 3) may have a more positive bias or inclination toward the software, these pools were used for recruitment for phases 2 and 3 to increase quality of feedback (i.e. for phase 2: high use = more familiarity with the system = higher quality feedback) and participation (i.e. for phase 3: enthusiasm for new software tools = higher likelihood of participation). Future studies employing more randomized selection of participants and larger numbers of participants could also (1) perform A/B testing on elements of the interface – particularly the display of patients lists and commonly confusing navigation items – to further evaluate intuitiveness versus adaptation/learning in the use of and development of proficiency with these aspects of the EMERSE system, or (2) compare EMERSE to other system(s) on the same tasks on the same dataset to provide a true comparison of the utility and usability of EMERSE versus other tools intended for the same tasks.

4.2 Conclusions

Overall, the EMERSE system performs very highly, was rated very positively by new and experienced users alike, and was met with great enthusiasm for its adoption in new sites.

In preparing for implementation of the system at new sites, key concerns lie in the realms of obtaining organizational approval and the ability to effectively export free-text note data from the site's EHR into the EMERSE system's database. While experienced users would unanimously recommend EMERSE to others and desire to see the continued expansion of the sources of documents indexed, they note that its use and utility is not as widespread and well-publicized as it could be. Participants expressed the belief that expansion of the document sources indexed would further increase the utility of the system in their work, and that better promotion and advertisement of the EMERSE system would facilitate the adoption of EMERSE by more users. Novice users noted in testing the system that navigation and affordances of the interface were not always completely intuitive, but believed they could quickly gain proficiency and expressed enthusiasm for the adoption of the tool in their work. Finally, experienced and novice users of EMERSE alike noted issues of clarity regarding which set of patients is currently selected and being searched through. Potential reasons for and solutions to this issue have been discussed. In summary, lessons learned from this study indicate the importance of clarity of navigation and apparentness of affordances in the usability of clinical documentation search tools. While small improvements can still be made in these areas, the EMERSE system and the results of this study present a template for using system colors and a highly intuitive user interface to create high-utility CIE tools, which are desperately needed in a field dominated by low-usability EHRs.

Bibliography

[1] S. Babbott, L. B. Manwell, R. Brown, E. Montague, E. Williams, M. Schwartz, E. Hess, and M. Linzer. Electronic medical records and physician stress in primary care: results from the MEMO Study. *Journal of the American Medical Informatics Association*, 21(e1):e100–e106, Feb. 2014.

[2] M. Burnett. Doing Inclusive Design: From GenderMag in the Trenches to Inclusive Mag in the Research Lab. *AVI*, 2020.

[3] J. P. Chin, V. A. Diehl, and K. L. Norman. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '88, pages 213–218, New York, NY, USA, May 1988. Association for Computing Machinery.

[4] J. W. Creswell. *Research design: qualitative, quantitative, and mixed methods approaches*. SAGE Publications, 4th ed edition, 2014.

[5] J. W. Creswell and C. N. Poth. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. SAGE Publications, Dec 2016. Google-Books-ID: DLbBDQAAQBAJ.

[6] J. Dabrowski, E. Letier, A. Perini, and A. Susi. Finding and Analyzing App Reviews Related to Specific Features: A Research Preview. In E. Knauss and M. Goedicke, editors, *Requirements Engineering: Foundation for Software Quality*, Lecture Notes in Computer Science, pages 183–189, Cham, 2019. Springer International Publishing.

[7] M. Design. Microsoft Inclusive Design Toolkit.

[8] M. W. Friedberg, P. G. Chen, K. R. Van Busum, F. Aunon, C. Pham, J. Caloyeras, S. Mattke, E. Pitchforth, D. D. Quigley, R. H. Brook, F. J. Crosson, and M. Tutty. Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy. *Rand Health Quarterly*, 3(4):1, Dec. 2014.

[9] C. Friedman and M. Rigby. Conceptualising and creating a global learning health system. *International Journal of Medical Informatics*, 82(4):e63–71, Apr 2013.

[10] C. P. Friedman, A. K. Wong, and D. Blumenthal. Achieving a nationwide learning health system. *Science Translational Medicine*, 2(57):57cm29, Nov 2010.

[11] D. A. Hanauer. EMERSE: The electronic medical record search engine. *AMIA Symposium*, page 941, 2006.

[12] D. A. Hanauer, J. S. Barnholtz-Sloan, M. F. Beno, G. Del Fiol, E. B. Durbin, O. Gologorskaya, D. Harris, B. Harnett, K. Kawamoto, B. May, E. Meeks, E. Pfaff, J. Weiss, and K. Zheng. Electronic medical record search engine (EMERSE): An information retrieval tool for supporting cancer research. *JCO Clinical Cancer Informatics*, pages 454–463, 2020.

[13] D. A. Hanauer, Q. Mei, J. Law, R. Khanna, and K. Zheng. 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). *Journal of Biomedical Informatics*, 55:290–300, 2015.

[14] D. A. Hanauer, D. T. Wu, L. Yang, Q. Mei, K. B. Murkowski-Steffy, V. Vydiswaran, and K. Zheng. Development and empirical user-centered evaluation of semantically-based query recommendation for an electronic health record search engine. *Journal of biomedical informatics*, 67:1–10, 2017.

[15] J. R. Hill, S. Visweswaran, X. Ning, and T. K. Schleyer. Use, impact, weaknesses, and advanced features of search functions for clinical use in electronic health records: A scoping review. *Applied Clinical Informatics*, 12(03):417–428, May 2021.

[16] A. K. Jha. The promise of electronic records: around the corner or down the road? *JAMA*, 306(8):880–881, Aug 2011.

[17] T. Johann, C. Stanik, A. M. A. B, and W. Maalej. SAFE: A Simple Approach for Feature Extraction from App Descriptions and App Reviews. In *2017 IEEE 25th International Requirements Engineering Conference (RE)*, pages 21–30, Sept. 2017. ISSN: 2332-6441.

[18] A. W. Kushniruk and V. L. Patel. Cognitive and usability engineering methods for the evaluation of clinical information systems. *Journal of Biomedical Informatics*, 37(1):56–76, Feb. 2004.

[19] Y. Lee, K. Kozar, and K. Larsen. The Technology Acceptance Model: Past, Present, and Future. *Technology*, 12, Jan. 2003.

[20] W. Maalej, Z. Kurtanović, H. Nabil, and C. Stanik. On the automatic classification of app reviews. *Requirements Engineering*, 21(3):311–331, Sept. 2016.

[21] W. Maalej and H. Nabil. Bug report, feature request, or simply praise? On automatically classifying app reviews. In *2015 IEEE 23rd International Requirements Engineering Conference (RE)*, pages 116–125, Aug. 2015. ISSN: 2332-6441.

[22] S. Magal and D. Mirchandani. Validation of the Technology Acceptance Model for Internet Tools. *Americas Conferences on Information Systems Proceedings*, page 7, 2001.

[23] M. Marathe and K. Toyama. Semi-Automated Coding for Qualitative Research: A User-Centered Inquiry and Initial Prototypes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Montreal QC Canada, Apr. 2018. ACM.

[24] D. W. Massaro. Review of The Psychology of Everyday Things. *The American Journal of Psychology*, 103(1):141–143, 1990. Publisher: University of Illinois Press.

[25] Microsoft. Microsoft Design.

[26] J. Nielsen. 10 usability heuristics for user interface design.

[27] J. Nielsen. Heuristic evaluation. In *Usability inspection methods*, pages 25–62. John Wiley & Sons, Inc., USA, June 1994.

[28] J. Nielsen. *Usability Engineering*. Morgan Kaufmann, Oct. 1994. Google-Books-ID: 95As2OF67f0C.

[29] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90*, pages 249–256, Seattle, Washington, United States, 1990. ACM Press.

[30] D. Norman. The psychology of everyday things. *American Journal of Psychology*, 1990.

[31] D. Norman. Emotion & Design: Attractive Things Work Better. *Interactions Magazine*, 9:36–42, July 2002.

[32] D. A. Norman, J. T. Stasko, J. Pierce, C. M. Potts, and C. Shaw. Design of everyday things. In *Design of Everyday Things*, 1988.

[33] H. Persson, H. Åhman, A. Yngling, and J. Gulliksen. Universal design, inclusive design, accessible design, design for all: different concepts—one goal? On the concept of accessibility—historical, methodological and philosophical aspects. *Universal Access in the Information Society*, 2014.

[34] L. Seyfried, D. A. Hanauer, D. Nease, R. Albeiruti, J. Kavanagh, and H. C. Kales. Enhanced identification of eligibility for depression research using an electronic medical record search engine. *International Journal of Medical Informatics*, 78(12):e13–e18, 2009.

[35] C. A. Sinsky, J. W. Beasley, G. E. Simmons, and R. J. Baron. Electronic Health Records: Design, Implementation, and Policy for Higher-Value Primary Care. *Annals of Internal Medicine*, 2014.

[36] P. E. University of Michigan. EMERSE documentation.

[37] P. E. University of Michigan. Publications using EMERSE.

[38] V. Venkatesh and F. D. Davis. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46(2):186–204, Feb. 2000. Publisher: INFORMS.

[39] Y. Wang, L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, S. Liu, Y. Zeng, S. Mehrabi, S. Sohn, and H. Liu. Clinical information extraction applications: A literature review. *Journal of Biomedical Informatics*, 77:34–49, Jan. 2018.

[40] L. Yang, Q. Mei, K. Zheng, and D. A. Hanauer. Query log analysis of an electronic health record search engine. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2011:915–924, 2011.

[41] P.-Y. Yen and S. Bakken. Review of health information technology usability study methodologies. *Journal of the American Medical Informatics Association*, 19(3):413–422, May 2012.

[42] K. Zheng, Y. Chen, J. Adler-Milstein, A. L. Rosenberg, D. T. Y. Wu, Q. Mei, and D. A. Hanauer. How do healthcare professionals personalize their software? a pilot exploration based on an electronic health records search engine. *Studies in Health Technology and Informatics*, 264:1408–1412, Aug 2019.

[43] K. Zheng, Q. Mei, and D. A. Hanauer. Collaborative search in electronic health records. *Journal of the American Medical Informatics Association*, 18(3):282–291, May 2011.

[44] H. Zhu, T. Gruber, and H. Dong. Value and Values in Inclusive Design. In *HCI*, 2020.

Appendix A

Phase 1 Protocol

A.1 Site Leadership Interviews: Protocol

1. What is your role at your institution?

2. How did you learn about EMERSE and became interested?

3. What are the audiences at your institution that you planned to serve using EMERSE?

4. Before EMERSE, did you have any free-text search system or capability (or NLP) to retrieve information from clinician notes or other types of free-text documents? *If yes, can you please describe:*

   (a) How well was it serving your users?

   (b) What type of issues did you have with the prior system or capability that prompted you to look for alternatives such as EMERSE?

   (c) What advantages did you see with EMERSE over what you had?

(d) What type of costs were associated with the prior system or capability?

(e) Does EMERSE replace, or augment, what you had?

5. Do you have any NLP capability at our institution and if so, how do you compare it to EMERSE?

6. What is the current status of EMERSE implementation supported through this NCI project?

7. To date, what is the most challenging aspect for implementing EMERSE at your institution?

   Probing Questions

   (a) Were there things that you feel you should have been done or done better to make the implementation a smoother process?

   (b) Were there things that you feel the Michigan team should have done or done better to make the implementation a smoother process?

   (c) Were there critical resources that should have be made available yet were not?

   (d) Besides what you mentioned above, are there other challenges that you have run into, or anticipate you will run into, when implementing EMERSE at your institution?

8. Were there things that you feel you did right which facilitated the implementation?

9. Were there things that you feel the Michigan team did right which facilitated the implementation?

10. If EMERSE has been rolled out to end users at your institution: what is the general feedback you have received from them about the system?

11. What challenges do you foresee with maintaining and supporting EMERSE in longer term?

12. Can you compare the EMERSE implementation to your implementation experience of other vendor products or open-source tools?

13. Anything else you want to tell us?

A.2 Technical Documentation

As this portion of the interview addresses some technical aspects of the implementation of the EMERSE system, technical documentation may be of interest to some. This documentation provides information on the system requirements for an installation of the EMERSE software as well as associated support systems and institutional and IT processes to facilitate implementation. Technical documentation can be viewed on the EMERSE Documentation website[36].

Appendix B

Phase 2 Protocol: Semi-Structured Interview

B.1 Overview

At baseline, the study will interview 20 veteran EMERSE users at the University of Michigan who have been using the system as a routine part of their job. These interviews will be semi-structured, and will include questions informed by technology acceptance theories. The purpose is to solicit feedback about the system's usefulness and ease of use, and to identify usability issues that may not be apparent to the development team. The interview protocol can be found in Appendix 1. Each interview will last 45 minutes to an hour; all will be conducted in private settings. Upon the participant's consent, the interview will be tape-recorded, and transcribed to facilitate subsequent qualitative analyses. No identifying information will be collected during the interviews. Any potential identifying information accidentally disclosed will be removed from the transcript. The audio tapes will be destroyed once they are transcribed.

Recruitment

The 20 veteran EMERSE users at the University of Michigan will be identified and recruited by the PI of the larger NCI-funded project, Dr. David Hanauer. These participants will be engaged users who have contributed to the development, testing, and evaluation of the EMERSE system in the past.

Purpose

To solicit feedback about EMERSE's usefulness and ease of use, and to identify usability issues that may not be apparent to the development team.

B.2 Interview Protocol

B.2.1 General Instructions

Avoid asking for information that would uniquely identify the interviewee.
A question may be skipped if the interviewee has adequately addressed it in an earlier part of the conversation.
A probing question may be skipped if the interviewee has adequately addressed it in an earlier part of the conversation.

B.2.2 Introduction

Purpose: To introduce the study.
Suggested time: 3 minutes

1. Introduce yourself.

2. Introduce the study:

```
"Thanks for your continued support of EMERSE. As you know,
```

```
the objective of this interview is to better understand issues
that you may have encountered when using EMERSE in your work,
in order to further improve the system.
This interview will take approximately 30 to 45 minutes. Your
participation and your responses will be treated confidentially.
All of our findings will be reported anonymously. Nothing that
you say will be traceable to you as an individual. We greatly
appreciate a recording of this interview for analysis purposes."
```

3. Hand out Informed Consent.

4. Answer any question the participant may have about the Informed Consent, and then have the participant sign it.

B.2.3 Descriptive & Background Questions

Purpose: Warm up questions to gather general facts about the interviewee and the work environment.

Suggested time: 5 minutes

```
"I'd like to start with some questions about your position
here and your general work setting."
```

1. What's your job role? (Q1)

   Note: While we have this information already, the goal of this question is to warm the participant up.

   Probing questions

   (a) What is your job title?

(b) What are your main job responsibilities?

(c) What kind of medical data do you frequently work with?

2. How long have you been working in this capacity? (Q2) *No probing questions. Let the participant speak.*

B.2.4 EMERSE Questions

Purpose: Questions to gather specific information about health IT's impact on workflow.

Suggested time: 20–30 minutes

1. How long have you been an EMERSE user? (Q3)

   No probing questions. Let the participant speak.

2. How often do you use the system? (Q4)

   No probing questions. Let the participant speak.

   Do not define time framing for the participant (e.g. how many times a day, a month). Let the participant decide how to report frequency.

3. How did you discover the system? (Q5)

   Probing questions

   (a) How did you retrieve information from electronic health records before you discovered EMERSE?

   (b) How was the experience like, i.e., retrieving information from electronic health records without the assistance of EMERSE?

4. In your current or previous work, have you used any other systems to help with retrieving information from electronic health records? (Q6)

   Probing questions

(a) If so, names of these systems and how do they compare to EMERSE

(b) Overall impression about these systems, in comparison with EMERSE

5. In general, how do you like EMERSE? (Q7)

   No probing questions. Let the participant speak.

6. What are things that you do not like about EMERSE? (Q8)

   No probing questions. Let the participant speak.

7. If we have a magic wand to change EMERSE any way we want, what would the first thing coming into your mind that should be changed? (Q9)

   Probing questions

   (a) Rationale for the change

   (b) What would the ideal system look like after the change is made?

8. Would you recommend EMERSE to other people who do similar work as yours? (Q10)

   No probing questions. Let the participant speak.

B.2.5 Additional Questions

Purpose: Questions to gather additional feedback.

Suggested time: 5–10 minutes

1. When you run into a problem using EMERSE, who do you go to for help? (Q11)

   No probing questions. If possible, ask the participant to provide specific names.

2. Overall, what do you think about the support you have received by the EMERSE team? (Q12)

   No probing questions. Let the participant speak.

B.2.6 Section 5: Wrap-Up

Purpose: To collect additional information that the participant may want to provide.

Suggested time: 5 minutes

1. Is there anything else that you'd like to share with us regarding your experience with using EMERSE? (Q13)

   No probing questions. Let the participant speak.

```
"Thank you very much for taking the time to participate in the
study. We appreciate it much your time and your help."
```

Appendix C

Phase 3 Protocol: EMERSE Workflow Scenarios

C.1 Overview

Following EMERSE's initial installation and functional testing, we will identify 5 prospective users at each collaborating site to participate in additional usability studies to reveal potential issues that may arise from the sites' local contexts. These users will be asked to review the user documentation, and then follow a structured test script to perform a set of simulated search tasks. They will then be asked to report their perceptions and experience with the system using the Questionnaire for User Interface Satisfaction (QUIS), a validated and widely used usability survey instrument (Appendix 2). They will also be asked to fill out a validated questionnaire instrument based on the technology acceptance model (Appendix 3), which assesses key determinants of technology acceptance behavior among prospective users. Each usability testing session, including the time needed to respond to the questionnaires, will last about an hour. No identifying information will be collected.

Recruitment

The 5 prospective users at each collaborating site will be identified and recruited by the site PIs. These participants are usually clinician scientists, clinical research coordinators, or healthcare administrators, who frequently perform medical chart review tasks that can be facilitated by the use of the EMERSE system.

C.2 Protocol

C.2.1 Background

You will be using a test system that contains 10,000 simulated patients and about 500,000 simulated clinical documents. All names, dates, documents, etc. are fake. A small handful of scanned documents are real but have been donated for use and have had all identifiers removed.

Details

You will be given three short scenarios to work through to familiarize yourself with the EMERSE system. You do not need to have prior experience with using EMERSE in order to complete these scenarios. Please follow the instructions in the order in which they are listed. The main task you are being asked to do is show in the Instructions column. Additional details to help understand what to do or the context for how EMERSE works can be found in the Explanation column. After you have worked through a scenario, feel free to try out other aspects/features of the system in any way you'd like–there is nothing you can "break". The three scenarios appear on the following pages.

C.2.2 Scenario 1

Table C.1: Scenario 1 workflow instructions as given to participant.

Step	**Instructions**	**Explanation**
1	Login	You will be given the URL, username, and password to use for logging in.
2	Attest to the use case *Quality Improvement*	An attestation is a way to record why you are using EMERSE for the particular session.
3	Enter the search term "seizures" and then use the *Find Patients* option to search across *All Local Patients* in the system	*Find Patients* will identify a set of patients that contain the term. You can apply the *Find Patients* function to an existing list of patients or to *All Local Patients* in the system. The default after logging in is that *All Local Patients* are automatically selected unless you change it. The word *Local* is used because a new *Network* feature will be coming out where you will be able to get a count of patients from other medical centers. Also note that EMERSE is very literal, so you if you search for "seizures" (with an "s" at the end) it will only find that specific variation. There are options to include other variations, which will come up later.
4	Determine how many patients mention the word "seizures"	You should see that there are 1,994 patients.

Continued on next page

Table C.1 – *Continued from previous page*

Step	Instructions	Explanation
5	Examine the list of 1,994 retrieved patients in more detail by moving it to a *Temporary Patient List*	A *Temporary Patient List* is a list of patients that is not saved between sessions. It is good for doing a quick review but where you should not be saving the results (such a "review preparatory to research"). *Temporary Patient Lists* can be converted to *Saved Patient Lists* when desired.
6	Assume that the first three patients are not relevant, so remove them from the list	This is done by clicking on the *Remove* link in the table for each patient.
7	Use the *Highlight Documents* option to see what it says about Heidi Kent in terms of where "seizures" is mentioned in the `Main EHR` category	Click on the *Highlight Documents* button towards the upper left of the screen. Heidi Kent should be the 6th patient down from the top. Click where it says "1 of 16" for her row and the `Main EHR` column. This means that 1 out of 16 documents mentions seizures. You can click again on the row with the snippet that mentions "seizures" to see the term in the context of the full document.
8	Look to see the context of the term "seizures" in Heidi Kent's radiology reports	This can be done simply by switching over to the `Radiology` tab.

C.2.3 Scenario 2

Table C.2: Scenario 2 workflow instructions as given to participant.

Step	**Instructions**	**Explanation**
1	Remove the "seizures" search term	Navigate back to *Terms* by clicking on the *Terms* button (upper left). A term can be removed several ways. One way is to click on the pencil icon and then click on *Remove*
2	Add the term "carpal tunnel syndrome"	You do not need to add double quotes around the phrase since multiple words in one row will be considered to be a quoted phrase (quotes will be added automatically).
3	Use the *Synonyms* feature to include all of the other possible terms **except** for the term "cts"	Click on the "Synonyms" button next to the phrase "carpal tunnel syndrome". To de-select the term "cts" click on it once. It should turn gray with a line through it. The other terms still highlighted in yellow will be added. Then click on the *Add Highlighted Terms* button.
4	Set the patients back to *All Local Patients*	Click on the *Patients* button and then the *All Local Patients* tab/header, and then select the Checkbox in the row in the table that says "All patients in the EMERSE system"

Continued on next page

Table C.2 – *Continued from previous page*

Step	**Instructions**	**Explanation**
5	Use the *Find Patients* feature to find patients that contain any of the terms related to "carpal tunnel syndrome" or the variations that were added.	When searching across *All Local Patients*, EMERSE treats terms with the same color as being separated by "OR" and terms with different colors to be separated by "AND". In this case, since the terms added were synonyms of "carpal tunnel syndrome" the system added them as the same color. The final result should be 164 patients.
6	Move the patients to a *Temporary Patient List*	
7	Convert the patient list to a *Saved Patient List*. Give it the name "Carpal Tunnel" and give it the Description "EMERSE testing".	A *Saved Patient List* can be shared with other users on your team and also supports *Comments* and *Tags*. A *Comment* is a brief note that you can make about the patient, and a *Tag* is a checkbox for the patient that you can use in any way you want. For example, you might want to Tag patients eligible for a study or that you have further questions about.
8	Go back to the *Terms* and *Clear All Terms*	Navigate to *Terms*, then click on *Clear/Delete*, then click on *Clear All Terms*
9	Add in a new term "tingling"	Navigate to *Manage Terms* and then add the new term.

Continued on next page

Table C.2 – *Continued from previous page*

Step	Instructions	Explanation
10	Use the *Highlight Documents* feature to find where "tingling" appears in the patient notes	
11	Add or remove *Tags* for some of the patients.	This *Patient List* is shared with multiple users who all can all change the status of *Tags*.
12	Add/edit *Comments* for some of the patients.	Similar to the *Tags* this *Patient List* is being shared with multiple users, so you may see comments from them as well. *Comments* are saved automatically once you click out of the text box.)

C.2.4 Scenario 3

Table C.3: Scenario 3 workflow instructions as given to participant.

Step	Instructions	Explanation
1	Select a *Saved Patient List* that has been shared to you called "EMERSE Shared List"	Navigate to *Saved Patient Lists* and find the list with the name "EMERSE Shared List" in the table. Click on it to select it. The list should contain 50 patients.

Continued on next page

Table C.3 – *Continued from previous page*

Step	**Instructions**	**Explanation**
2	Select a *Term Bundle (Saved Terms)* called "EMERSE Testing Bundle 1"	Navigate to *Terms* and then *Saved Terms*. Then choose the radio button labeled *All Available*. Click on the row with the name "EMERSE Testing Bundle 1". This should select the list of terms which should now be displayed near the top of the screen, next to word *Terms*.
3	View the Overview of the patients by clicking on the *Highlight Documents* button.	This should show you the high level *Overview* of which documents had the term(s) of interest. Some of the cells in the *Overview* should have numbers in them, like "1 of 14", which means that 1 document out of 14 had a term of interest.
4	Click on the *Mosaic* icon which is above and on the right side of the *Overview* table.	This should switch to a view where each of the colored terms appears as a color in the table, allowing you to identify which term(s) are present just by their colors.
5	Locate the patient that has the terms "coronary artery disease" (blue) and "triglycerides" (red) appearing in their *Radiology Notes* and click on that cell.	This should be for patient "Brielle Kelley" and by clicking you should be able to identify two notes, one with the term "triglycerides" and another with the term "coronary artery disease".
6	Logout of the system	Click on your username in the upper right to access the menu and choose Logout.

C.3 Surveys

C.3.1 QUIS

The QUIS survey is a validated measure of user interface satisfaction [3].

Table C.4: QUIS Survey for Phase 3

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
Overall Reaction to Software											
terrible	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	wonderful
difficult	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	easy
frustrating	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	satisfying
dull	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	stimulating
rigid	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	flexible
Screen											
characters were: hard to read	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	easy to read
highlighting simplified task: not at all	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	very much
organization of information was: confusing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	very clear
sequence of screens was: confusing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	very clear

Continued on next page

Table C.4 – *Continued from previous page*

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
Terminology & System Information											
Use of terms was: inconsistent	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	conssitent
terminology was related to task: never	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	always
position of messages was: inconsistent	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	consistent
messages prompting for input: confusing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	clear
computer keeps you informed about what it is doing: never	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	always
error messages: unhelpful	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	helpful
Learning											
learning to operate system: difficult	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	easy

Continued on next page

Table C.4 – *Continued from previous page*

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
exploring new features by trial and error:											
difficult	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	easy
remembering names and use of commands:											
difficult	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	easy
tasks can be performed in a straight-forward manner:											
never	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	always
help messages:											
unhelpful	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	helpful
supplemental reference materials:											
confusing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	clear
System Capabilities											
system speed:											
too slow	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	fast enough
search results quality:											
poor	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	good
correcting your mistakes:											
difficult	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	easy

Continued on next page

Table C.4 – *Continued from previous page*

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
experienced and inexperienced users' needs are taken into consideration:											
never	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	always
Usability & UI											
use of colors and graphics:											
poor	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	good
system feedback:											
poor	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	good
system response to errors:											
awkward	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	gracious
system messages and reports:											
poor	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	good
system clutter and UI "noise":											
poor	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	good

C.3.2 TAM

The TAM survey is based on the technology acceptance model and is validated in a number of studies[22, 19, 38].

Table C.5: TAM Survey for Phase 3

Question	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
Using this system in my job would enable me to accomplish tasks more quickly.	☐	☐	☐	☐	☐
Using this system would improve my job performance.	☐	☐	☐	☐	☐
Using this system would enhance my effectiveness on the job.	☐	☐	☐	☐	☐
Using this system would make it easier to do my job.	☐	☐	☐	☐	☐
I would find this system useful in my job.	☐	☐	☐	☐	☐
Learning to use this system would be easy for me.	☐	☐	☐	☐	☐
I would find it easy to get this system to do what I want it to do.	☐	☐	☐	☐	☐

Continued on next page

Table C.5 – *Continued from previous page*

Question	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
My interaction with this system would be clear and understandable.	☐	☐	☐	☐	☐
It would be easy for me to become skillful at using this system.	☐	☐	☐	☐	☐
I would find this system easy to use.	☐	☐	☐	☐	☐

Appendix D

Open Source Code

D.1 Repositories

Code for the transcription process in this project is open source and can be found at `https://github.com/colby-reyes/EMERSE`. The EMERSE system is also open source. Its repository can be found here: `https://github.com/project-emerse`. Notebook code, referenced above, for "SAFE" (semi-automated feature extraction) method-based semi-automated qualitative coding can be found at `https://github.com/colby-reyes/EMERSE/blob/main/EMERSE_FeatureExtraction.ipynb`

Appendix E

EMERSE Information

E.1 Learn More About EMERSE

Additional feature descriptions and screenshots of the EMERSE system can be found at `https://project-emerse.org/features.html`

E.2 Contact

Though the EMERSE project is open-source, the projects here are private since we want to track the usage of it since it is grant-funded. If you want accesss, send the EMERSE team a message at EMERSE-team@umich.edu.

Appendix F

Additional Tables and Figures

F.1 Description of Usability Heuristics

#	**Heuristic**	**Description**
1	Visibility of system status	The design should always keep users informed about what is going on, through appropriate feedback within a reasonable amount of time.
2	Match between system and the real world	The design should speak the users' language. Use words, phrases, and concepts familiar to the user, rather than internal jargon. Follow real-world conventions, making information appear in a natural and logical order.
3	User control and freedom	Users often perform actions by mistake. They need a clearly marked "emergency exit" to leave the unwanted action without having to go through an extended process.
4	Consistency and Standards	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform and industry conventions.
5	Error prevention	Good error messages are important, but the best

F.2 Quantitative Analysis: Extended Tables

Table F.2: Full table of user testing analysis.
Note: Error, Mistake, Help, and Search/Delay categories are not mutually exclusive

Step	**Instructions**	**Error**	**Temp. Mistake**	**Help**	**Search**	**Delay**	**No Issue**
	Scenario 1						
1	Login	0%	6.25%	0%	0%	0%	93.75%
2	Attest to the use case "Quality Improvement"	6.25%	0%	6.25%	0%	0%	93.75%
3	Enter the search term "seizures" and then use the Find Patients option to search across All Local Patients in the system	6.25%	6.25%	6.25%	12.5%	12.5%	81.25%
4	Determine how many patients mention the word "seizures"	0%	0%	0%	0%	0%	100%
5	Examine the list of 1,994 retrieved patients in more detail by moving it to a Temporary Patient List	0%	0%	0%	0%	0%	100%
6	Assume that the first three patients are not relevant, so remove them from the list	0%	6.25%	0%	6.25%	0%	93.75%

Continued on next page

Table F.2 – *Continued from previous page*

Step	**Instructions**	**Error**	**Temp. Mistake**	**Help**	**Search**	**Delay**	**No Issue**
7	Use the Highlight Documents option to see what it says about Heidi Kent in terms of where "seizures" is mentioned in the "Main EHR" category	0%	6.25%	12.5%	31.25%	31.25%	56.25%
8	Look to see the context of the term "seizures" in Heidi Kent's radiology reports	0%	0%	0%	0%	0%	100%
	Scenario 2						
1	Remove the "seizures" search term	0%	6.25%	0%	25%	25%	75%
2	Add the term "carpal tunnel syndrome"	0%	0%	0%	0%	0%	100%
3	Use the Synonyms feature to include all of the other possible terms except for the term "cts"	0%	12.5%	0%	0%	6.25%	87.5%
4	Set the patients back to All Local Patients	6.25%	0%	12.5%	18.75%	18.75%	0%

Continued on next page

Table F.2 – *Continued from previous page*

Step	Instructions	Error	Temp. Mistake	Help	Search	Delay	No Issue
5	Use the Find Patients feature to find patients that contain any of the terms related to "carpal tunnel syndrome" or the variations that were added.	0%	0%	0%	0%	6.25%	93.75%
6	Move the patients to a Temporary Patient List	0%	0%	0%	0%	0%	100%
7	Convert the patient list to a Saved Patient List. Give it the name "Carpal Tunnel" and give it the Description "EMERSE testing".	0%	0%	0%	0%	0%	100%
8	Go back to the Terms and Clear All Terms	0%	6.25%	0%	6.25%	6.25%	93.75%
9	Add in a new term "tingling"	0%	0%	6.25%	0%	0%	93.75%
10	Use the Highlight Documents feature to find where "tingling" appears in the patient notes	12.5%	0%	6.25%	0%	0%	81.25%
11	Add or remove Tags for some of the patients.	0%	12.5%	6.25%	12.5%	12.5%	75%

Continued on next page

Table F.2 – *Continued from previous page*

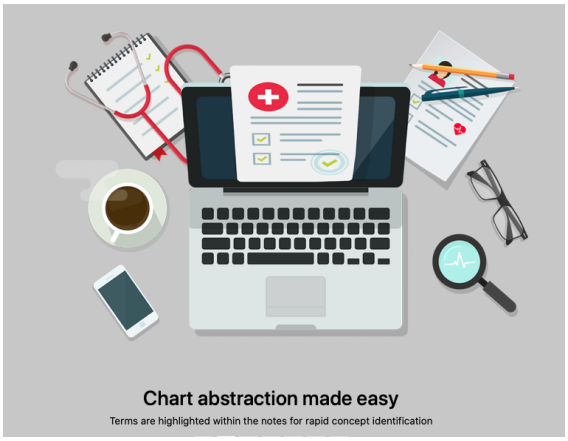
Step	**Instructions**	**Error**	**Temp. Mistake**	**Help**	**Search**	**Delay**	**No Issue**
12	Add/edit Comments for some of the patients.	0%	0%	0%	0%	0%	100%
	Scenario 3						
1	Select a Saved Patient List that has been shared to you called "EMERSE Shared List"	0%	0%	6.25%	6.25%	6.25%	93.75%
2	Select a Term Bundle (Saved Terms) called "EMERSE Testing Bundle 1"	6.25%	12.5%	6.25%	25%	18.75%	68.75%
3	View the Overview of the patients by clicking on the Highlight Documents button.	6.25%	6.25%	6.25%	6.25%	12.5%	75%
4	Click on the Mosaic icon which is above and on the right side of the Overview table.	0%	6.25%	0%	0%	6.25%	93.75%

Continued on next page

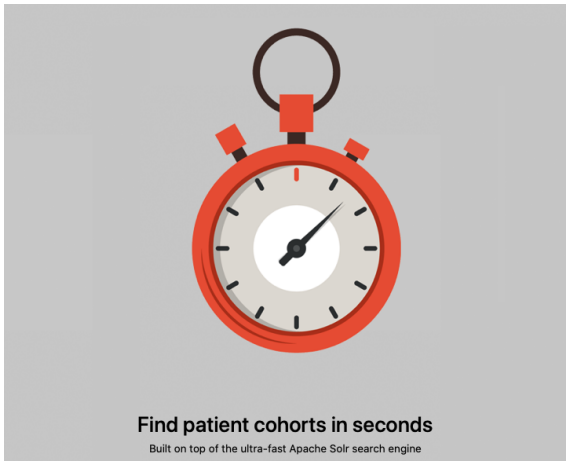
Table F.2 – *Continued from previous page*

Step	**Instructions**	**Error**	**Temp. Mistake**	**Help**	**Search**	**Delay**	**No Issue**
5	Locate the patient that has the terms "coronary artery disease" (blue) and "triglycerides" (red) appearing in their Radiology Notes and click on that cell.	0%	6.25%	0%	12.5%	18.75%	81.25%
6	Logout of the system	0%	0%	0%	0%	0%	100%

F.3 Qualitative Analysis: Charts and Figures



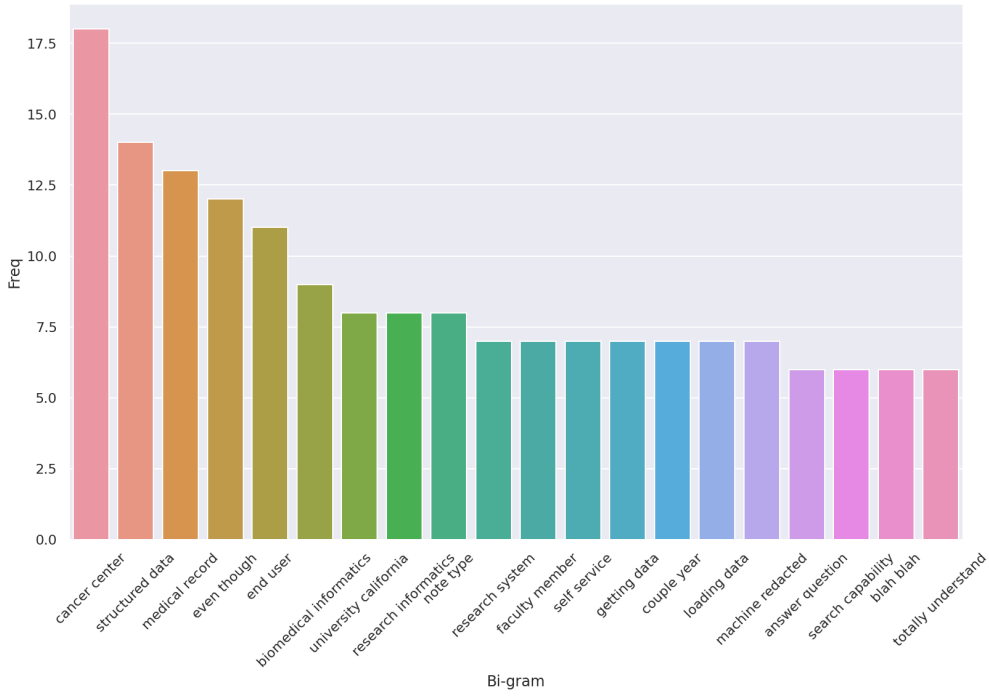
(a) Banner on using EMERSE for chart abstraction



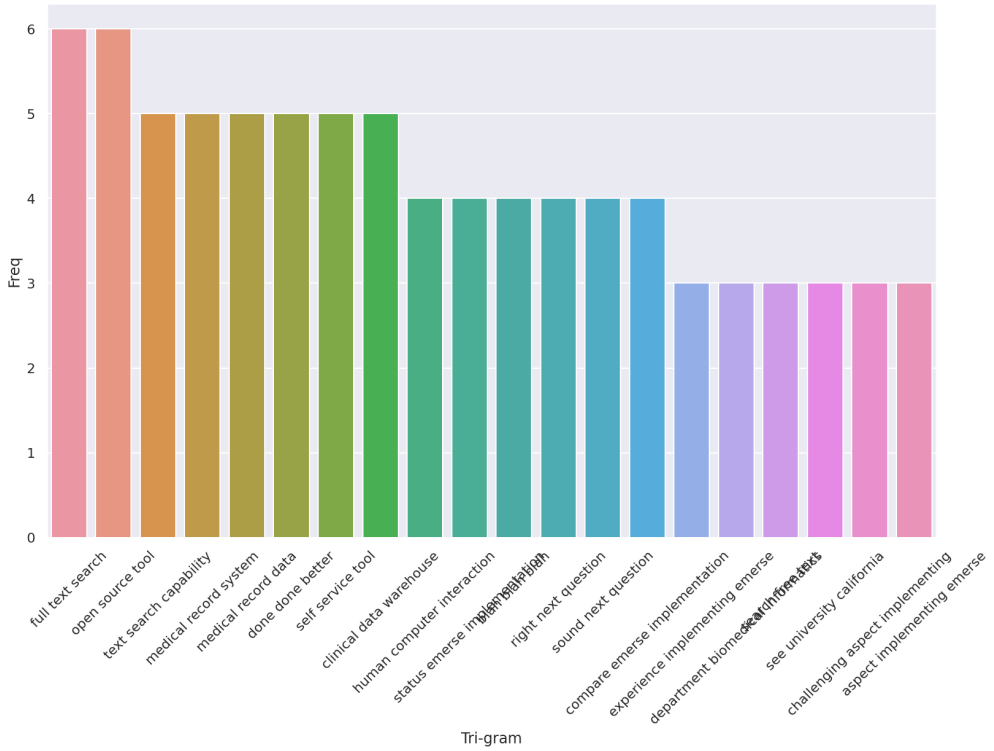
(b) Banner advertising the speed of EMERSE

(c) Banner promoting EMERSE as a data tool

Figure F.1: Banners from the EMERSE site

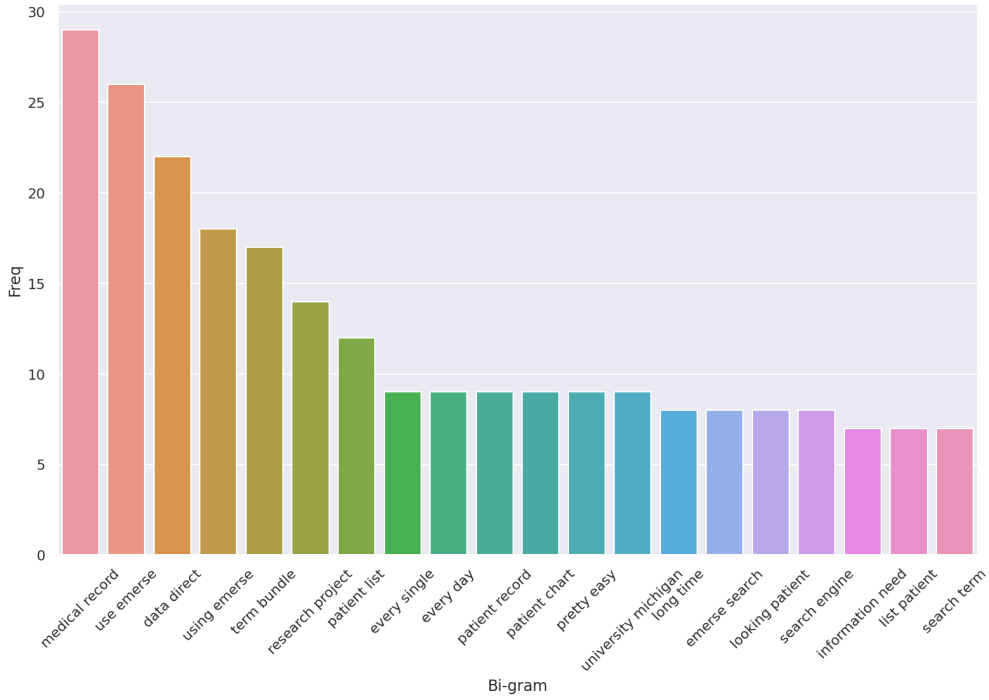


(a) Top bi-grams for Phase 1 interviews

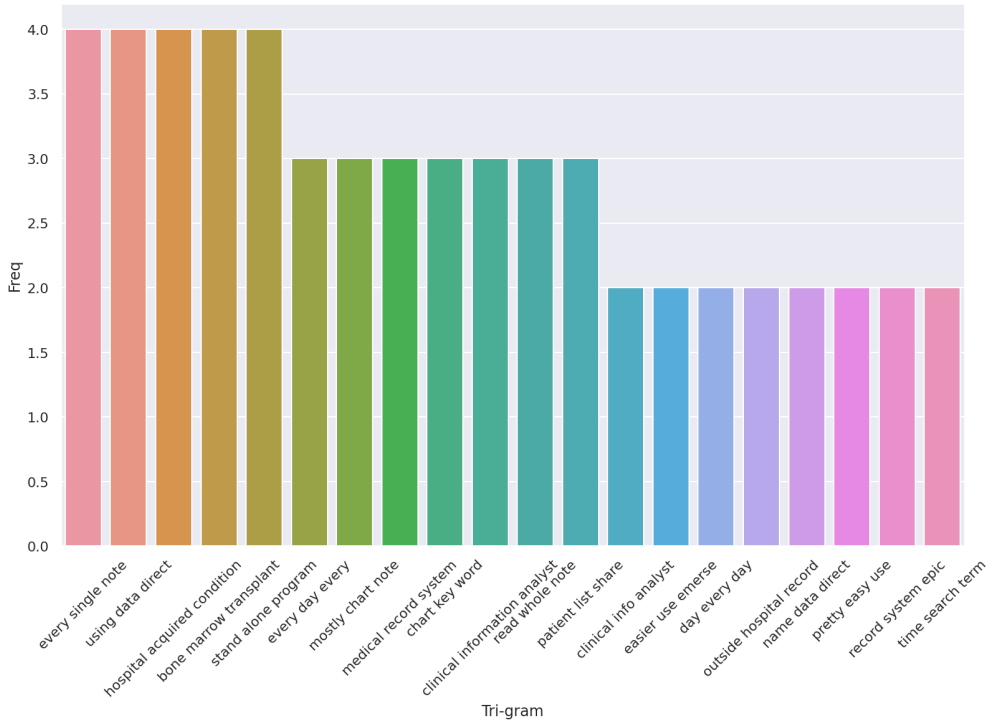


(b) Top tri-grams for Phase 1 interviews

Figure F.2: N-grams extracted from Phase 1 Transcripts

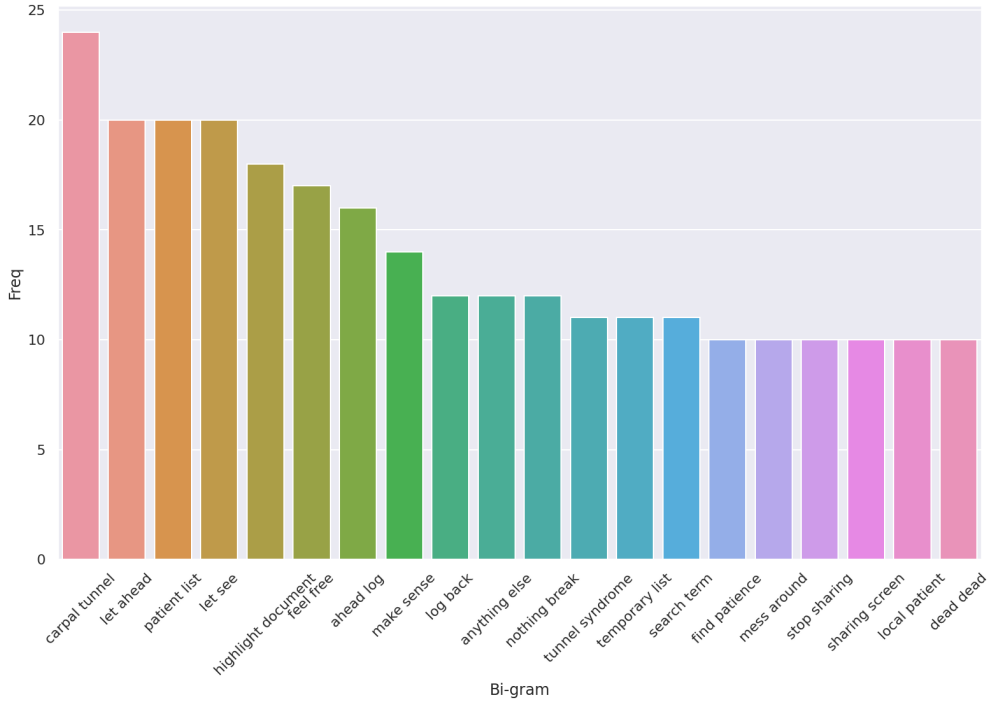


(a) Top bi-grams for Phase 2 interviews

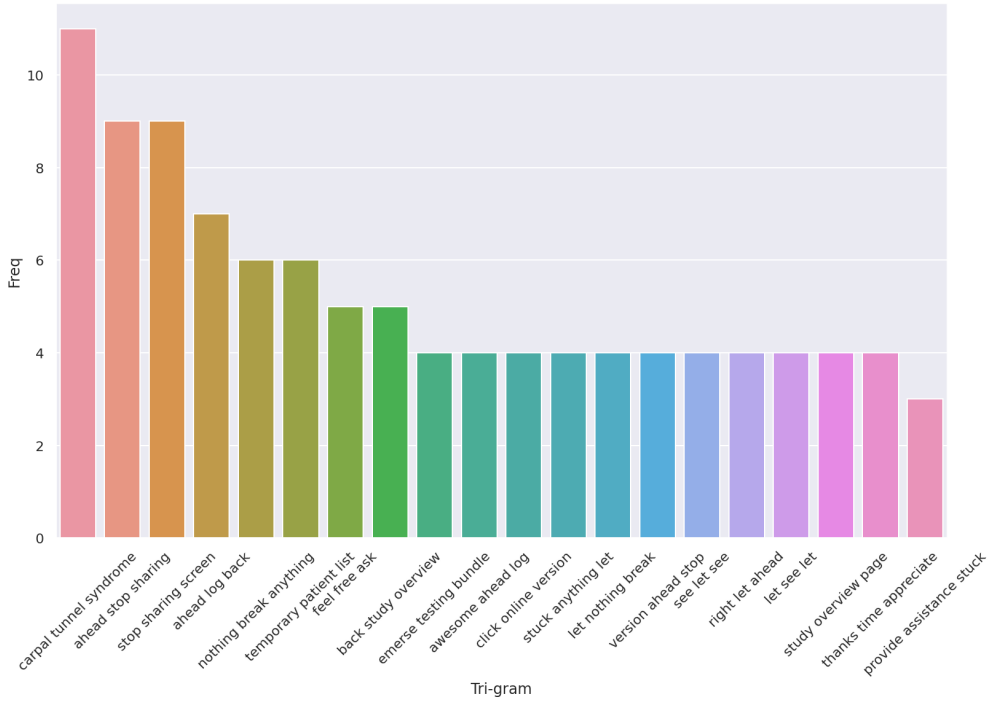


(b) Top tri-grams for Phase 2 interviews

Figure F.3: N-grams extracted from Phase 2 Transcripts



(a) Top bi-grams for Phase 3 interviews



(b) Top tri-grams for Phase 3 interviews

Figure F.4: N-grams extracted from Phase 3 Transcripts

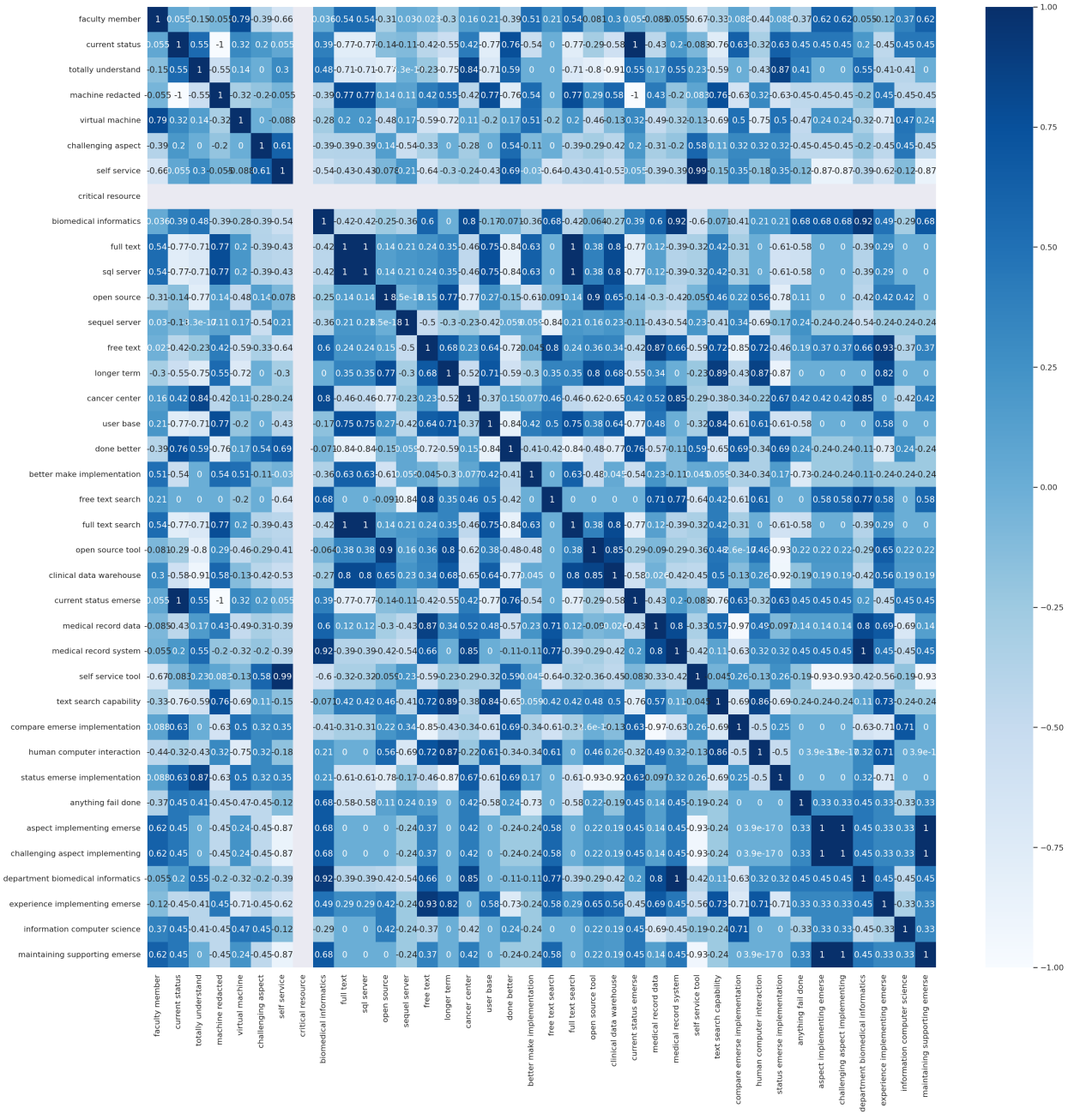


Figure F.5: Correlation of n-grams in Phase 1 Interviews

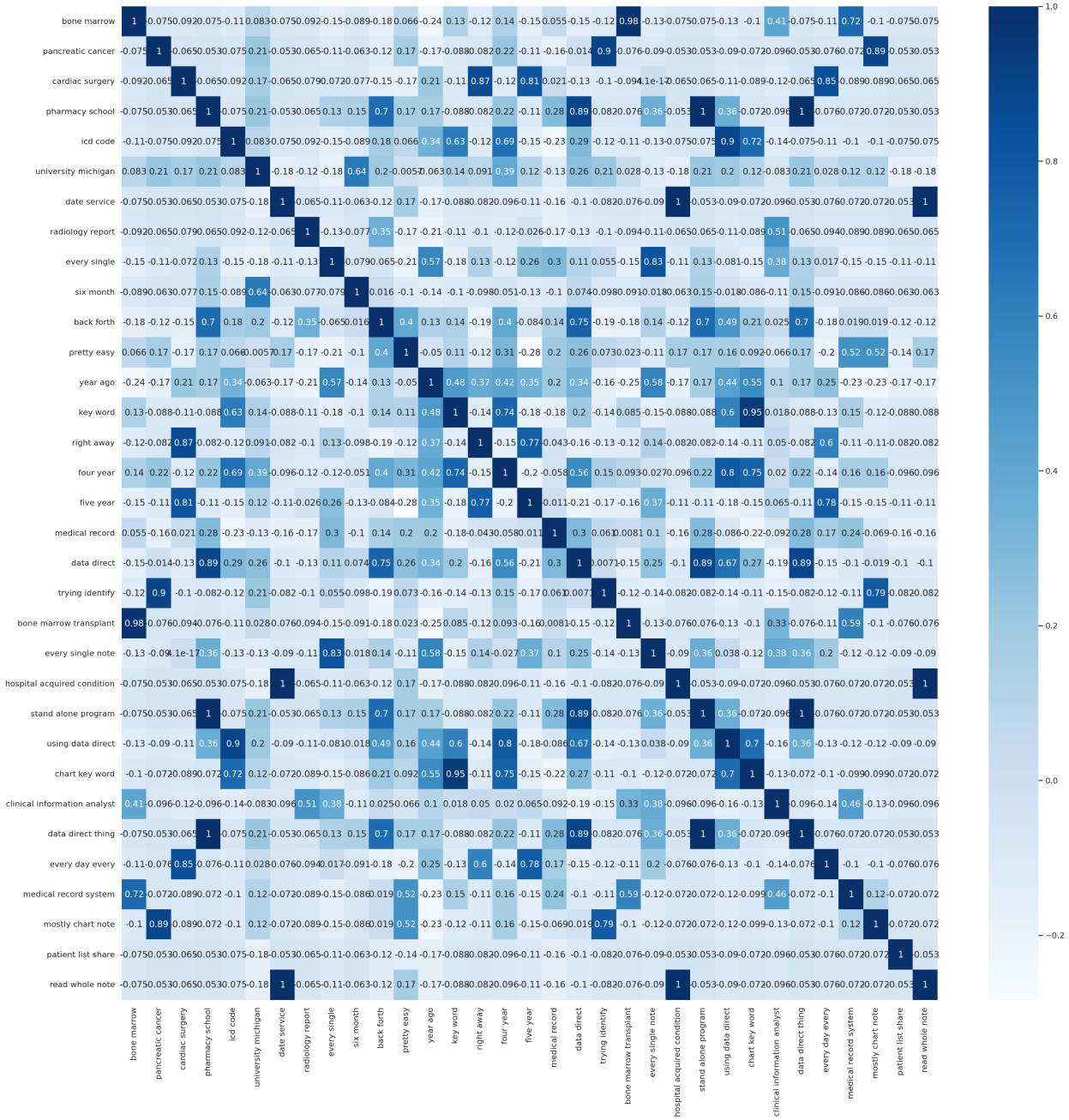


Figure F.6: Phase 2 text feature co-occurrences and correlations

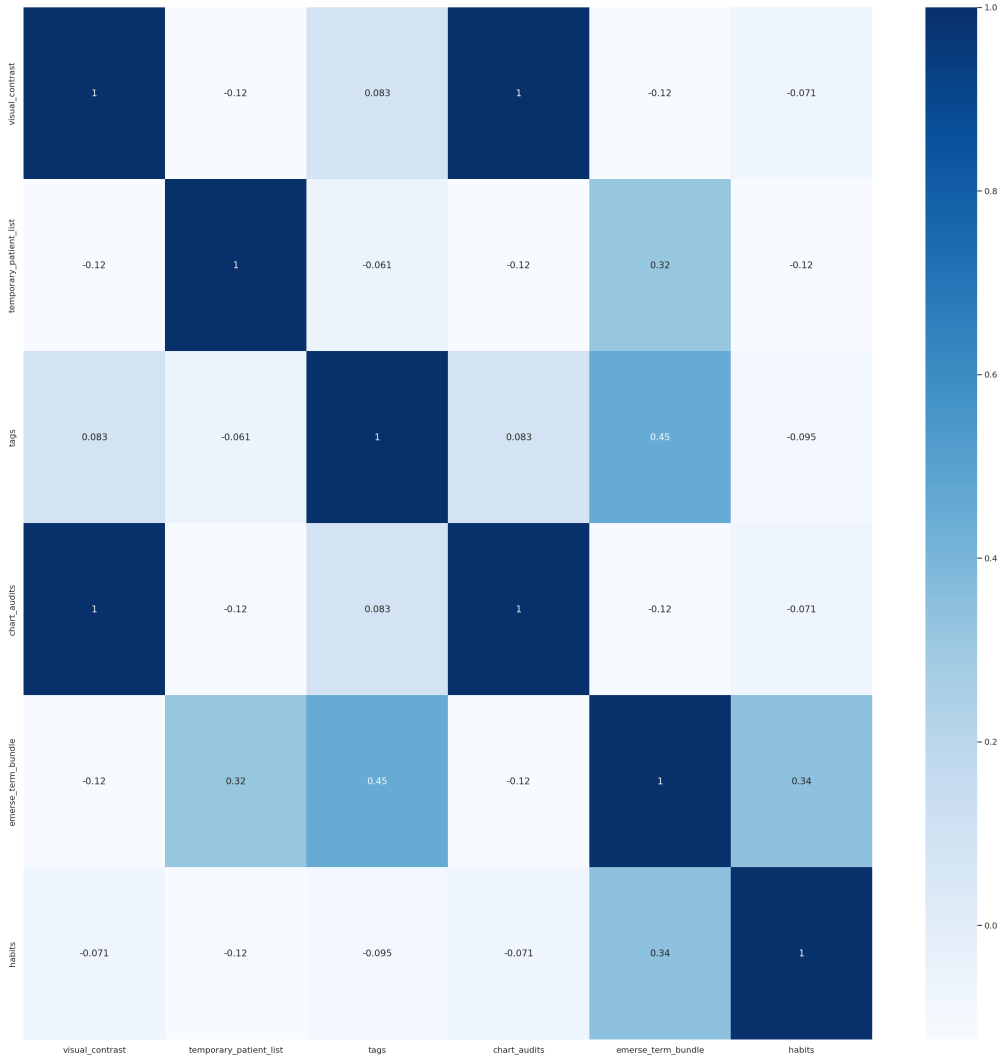


Figure F.7: Correlations for manually categorized n-grams in Phase 3