EMERSE

The Electronic Medical Record Search Engine

User Documentation

Updated March 24, 2006

This document contains confidential and proprietary information.
DO NOT distribute beyond the confines of the UM Health System.

ACKNOWLEDGEMENTS

EMERSE was developed with funding provided by the University of Michigan Comprehensive Cancer Center as part of the Comprehensive Cancer Center Bioinformatics Core.

The following individuals, listed in alphabetical order were instrumental in assisting with the development, funding, review, and deployment of EMERSE.

Douglas Blayney, MD	Comprehensive Cancer Center
William F. Bria, MD	Internal Medicine
Cynthia Carter	MCIT
Arul M. Chinnaiyan, MD, PhD	Pathology
Erdwing Coronado	MCIT
Kathy Davis	Pathology Data Systems
Douglas F. Gibbs, PhD	Pathology Bioinformatics
Erika Keith	MCIT
Alan Mahcinski	Pathology Data Systems
Vasudeva Mahavishnu	Cancer Center Bioinformatics Core
Rita Mitchell	Office of Clinical Affairs, CIDSS
Gary Mull	MCIT
Kurt Riegel	Compliance Office
Diane Shaltis	CACR
Kevin Smith, MS	Cancer Center Bioinformatics Core
Marcy Waldinger, MHSA	Comprehensive Cancer Center
Max Wicha, MD	Comprehensive Cancer Center

CONTENTS OF THIS DOCUMENT

QUICK START GUIDE

1. Log into EMERSE
2. Click on the “Enter CPIs” button
3. Type or paste in a list of CPI numbers, one per line
4. Click “Submit” – EMERSE will scan CareWeb to identify the patients
    a. You can give the patient list a name and save it for future use
5. Click on the “Search Terms” button
6. Type in a list of terms you wish to look for, separated by spaces
    a. EMERSE will search for all terms
    b. Phrases made of more than one word should be in double quotes
    c. See help section for more details about searching
7. Click “Submit” – EMERSE will acknowledge your selection and check for spelling errors
8. Click on the “Full Search” button – a table will slowly appear with results
    a. Each row represents a patient
    b. Each column represents a type of document
    c. Numbers in each cell represent the total number of documents with a “hit” over the total number of documents a patient has for that category
    d. Color codes are also indicative of the number of documents with a “hit”
9. Click on a cell in the table to drill down and see the summaries
    a. For all documents except the PSL, each row represents an individual document
    b. If the summary is blank, not hits for that document were found
    c. If the summary has text then at least on hit was found
10. Click on a row to drill down further and view the specific note with any hits in the text highlighted

Additional tips:

- Read the documentation about “Setting Up Search Terms In EMERSE” to utilize the full power of the tool’s searching capabilities
- Always think about how a concept might be worded in the note – use all abbreviations and synonyms
- The “Alternatives” link is a way to look for other words in the documents which are similar to your search terms that may represent spelling errors.
- “Bundles” are groups of search terms – these can be managed and shared with other users and are ideal for collaborative work as well as for lists containing many terms
- When in doubt, ask for help – contact David Hanauer

EMERSE INTRODUCTION

EMERSE is a powerful tool to help you search through the electronic medical record and find things quickly.

EMERSE:

- Is the Electronic Medical Record Search Engine
- Works on many documents in CareWeb including DMIs (Dictated Medical Information), Rad/Nuc Reports, Pathology Reports, and the PSL (Problem Summary List)
- Is a fast and efficient way to find text buried in the reports
- Is very useful for research and data abstraction
- Is web-based and has been tested on both Macintosh and Windows computers

EMERSE is similar to other search engines such as Google, but with notable differences. The most obvious difference is the interface. EMERSE was designed with the medical record in mind and organizes documents by type and even allows sorting documents based on dates. Another important difference is the rationale for using either search tool. Google is fantastic for trying to locate the "best" document that meets a user's criteria and ranks pages based on a proprietary and secret measure of relevance. EMERSE, however, is designed to find the *exact* information a user is seeking and does not perform any type of ranking. Google is great for finding the "best" piece of literature supporting a link between longevity and a low-fat diet and EMERSE is great for finding if your patient's specific electronic chart ever mentions the drug Lovastatin. For most practical purposes you wouldn't want to find the "best" patient or "best" document for which Lovasatin was mentioned, you just want to find the precise answer for your patient.

EMERSE and Google share a few other similarities which, along with other notable differences, are highlighted in Table 1 on the following page.

Searching medical text can be tricky and there are many nuances for which you should be aware. Furthermore, setting up your search parameters to find precisely what you're looking for takes some skill and a lot of forethought. This is true for any search engine, whether you use EMERSE or Google. This document will provide a brief introduction to the complexity of medical data and will also provide a tutorial on how to use EMERSE

EMERSE	Google
Does not index all data before a search is conducted; searches can sometimes take a little while	Indexes all data before searches are conducted; searches are virtually instantaneous
All users actions are logged in an audit trail	Allows for “anonymous” searching
Is secure and HIPAA compliant	Is not HIPAA compliant
Is designed to work specifically with medical documents	Is designed for any type of document, but doesn’t present the information in ways useful for medical documentation
Text searches can be performed for parts of a word (e.g. it can find “ear” in the word “heart”)	Searches only on distinct “words”, meaning that the text must be separated by white space on both sides.
Offers advanced features for improving searching capabilities	Offers advanced features for improving searching capabilities
Can suggest spelling corrections for search terms that might have been misspelled.	Can suggest spelling corrections for search terms that might have been misspelled.

Table 1. Similarities and differences between EMERSE and Google.

COMPLEXITY OF MEDICAL TEXT

Searching medical text can be complicated. There are many ways in which a health care provider can document the same concept. One must also consider the possible use of ambiguous abbreviations and other factors which can make teasing out the desired result an arduous task.

Many synonyms and abbreviations can exist for the same concept:

Carcinoma = Cancer = Neoplasm = Tumor

Zithromax = Zmax = Z-pak = Azithromycin

Heart attack = Myocardial infarction = MI

Breast Carcinoma = Breast Cancer = Breast Ca = BR CA = BRCA

White Blood Cell Count = White Count = WBC = WBC Count = Leukocyte Count

Abbreviations can be ambiguous:

Ca = calcium
CA = cancer
CA = California

Mg = Magnesium
mg = milligrams
MG = myasthenia gravis

ALL = acute lymphoblastic leukemia
ALL = allergies

Words can be adjacent, without intervening white space:

Cervical canal was normal [Cervical canal was normal]
Cervical canot identified [Cervical cancer not identified]

Case-sensitivity sometimes matters:

Did all of the patients in clinic with ALL arrive on time?

The renal patient with a high BUN can eat the bun but not the burger.

He had FROM of his hips bilaterally and could walk from the door to the chair.

Her DIP joint is swollen but here urine dip was normal.

CODED VERSUS FREE TEXT

Virtually all of the patient related data is in electronic form in our electronic medical record. In principle, it would be great it we could simply ask (aka "query") the database for the specific information of interest: *Please tell me which patients in my study had a post-surgical infection caused by MRSA.* If all of the information was coded, meaning a number was assigned to every concept, research would be simple; we could ask the computer to look through its database for patients that had a code for both *post-surgical infection* and *MRSA*. Codes make finding things easy because it standardizes the way in which a concept can be represented. Codes don't always have to be used for this: if standardized terminology is used (e.g. if MRSA is always written as "MRSA" or "Methicillin Resistant Staphlocococcus Aureus" such a search would at least, in principle, be relatively easy.

But our medical record stores almost all of its information in free-text. The "free" means that the users (in this case the health care providers) are free to say something anyway they want—and users have found many ways to say essentially the same concept. The following table shows just how varied things can be. An analysis of the problem summary list was conducted and some, but not all, diagnoses entered related to "strep throat" were collated in Table 2, below. You can see the large amount of variation. Some are clearly definitive diagnoses, some are less definitive based on the modifiers used ("probable", "possible", etc.) Others use wording that makes it unclear whether the diagnosis was even made or if it was simply ruled out ("rule out strep").

DEFINITIVE STREP DIAGNOSES Acute pharyngeal tonsillitis, secondary to strep Acute strep pharyngeal tonsilitis Acute strep pharyngitis Acute strep pharyngitis Acute strep pharyngotonsillitis Acute strep pharyngotonsillitis, counter positive Acute strep throat Acute Strep-positive pharyngeal tonsillitis Acute streptococcal pharyngotonsillitis Acute streptococcal tonsilitis. Acute streptococcaltonsillitis/ pharyngitis Group A beta strep exudated tonsillitis Group A strep infection Group A Strep pharyngitis Group A strep pharyngitis Group A streptococcal Pharyngitis Positive strep screen Scarlet fever, strep positive Strep throat Strep throat (Strep pharyngitis) Strep throat infection Streptococcal pharyngitis Streptococcus throat **POSSIBLE STREP DIAGNOSES** Empiric strep throat Exposure to Strep Exudative tonsillitis, suspect Strep Pharyngitis, strep probable. Phone report of positive strep screen Possible strep pharyngitis possible strep throat Presumed strep Presumed Streptococcal pharyngitis. Probable Strep pharyngitis Probable Strep throat symptoms of Strep with fever	**NEGATIVE STREP DIAGNOSES** Non-Strep pharyngitis, 03/22/2006 Pharyngitis non-streptococcal, rapid strep screen negative Sore throat strep negative Strep negative pharyngitis Viral upper respiratory infection with sore throat, rapid strep test negative **RULE OUT STREP** Acute pharyngeal tonsillitis, rule out Streptococcus Acute pharyngitis (rule out Strep) Acute pharyngitis, rule out strep Acute pharyngitis, rule out streptococcus Acute pharyngotonsillitis, rule out Strep Cervical lymphadenopathy, rule out strep Pharyngitis, rule out Strep Rule out recurrent strep Rule out strep Rule out Strep pharyngitis Rule out streptococcus pharyngotonsillitis with conjunctivitis There is a pharyngitis, rule out strep. **HISTORY OF STREP** Followup Strep pharyngitis history of multiple Strep throats Resolved strep pharyngitis **SPELLING ERRORS** Probable Strep throa sore throat with weakly positive strept screen Strep Pharengitis strept pharyngitis

Table 2. A subset of the diagnoses typed into the problem summary list dealing with strep throat. Trying to automatically extract concepts from the medical record can be difficult when so much variation is possible.

If this much diversity exists for a concept as simple and straightforward as *strep throat* (caused by Group A Strep) imagine the variation that may exist for complex medical conditions. Additionally, the strep terms in the above table don't include other types of "strep" such as "Group B Strep", or strep infections causing impetigo, or the "flesh-eating" strep.

ANALYSIS OF A SINGLE MEDICAL DOCUMENT

It is very hard to pull out all concepts from the medical record and store them in a database. What is considered a unique "concept" may vary from person to person, but even a short and simple note can be information dense. The following is an actual (de-identified) note from the medical record.

```
16 yo here with mom.

3 days frontal HA and fever. No ST, No cough, No RN, mild
abd pain,
sl dizzy today.  Perfectly well when she has has motrin.

No illnesses in years

T 98.8
Wt 165

TM perfect. Throat nl Chest clear. Turbs pink with exudate.

Imp URI. Course reviewed
```

Figure 1. An actual, brief medical note from the medical record. Although very short it contains over 20 concepts that could potentially be extracted.

At least 21 concepts can be extracted from this brief note, as shown below in Table 3. Additionally, not all concepts include other modifiers that might be important such as duration, severity, timing, etc.

Age: 16 years old **Gender**: Female **Family history**: Mother is currently alive. **History of present illness**: • Frontal headache (duration = 3 days) • Fever (duration = 3 days, maximum temperature not specified) • Sore throat not present • Cough not present • Runny nose not present • Abdominal pain present (severity = mild) • Dizziness present (duration = 1 day, severity = slight) • All symptoms relieved by ibuprofen	**Past medical history**: No significant illness in many years (exact number not specified) **Physical exam:** • Temperature: 98.8 (Fahrenheit implied, but not units given) • Weight: 165 (Pounds implied, but units not given) • Tympanic membranes normal • Throat normal • Chest clear • Turbinates pink • Turbinates with exudates present **Impression:** Upper Respiratory Infection **Plan:** Course reviewed (details not specified)

Table 3. Some of the concepts which can be extract from the short note in Figure 1.

Furthermore, the document uses 13 abbreviations, not all of which would be considered standard (Table 4).

Abbreviation	Meaning
yo	years old
HA	headache
ST	sore throat
RN	runny nose
abd	abdominal
sl	slightly
T	temperature
Wt	Weight
TM	tympanic membranes
nl	normal
Turbs	turbinates
Imp	Impression
URI	Upper Respiratory Infection

Table 4. The 13 abbreviations used in the short note from Figure 1.

This note demonstrates the trouble of searching free-text documents—it's hard to know how someone chose to express an idea, concept, condition, finding, diagnosis, plan, etc. Nevertheless, with EMERSE you should still mostly be able to find what you need as long as you spend enough time thinking about what you're looking for.

ERRORS INTRODUCED INTO THE MEDICAL RECORD

Using EMERSE can never guarantee that you will find everything that you're looking for—it depends on how someone worded something. To understand why some errors exist, it's important to understand how information is entered in the system. Some health care providers type in the notes themselves (which are subject to typographical errors). However, the majority of notes are entered by transcriptionists who listen to dictated notes and then try to reproduce, as best as possible, what they hear. Sometimes the transcriptionists make errors because they don't have enough medical knowledge to understand the nonsense they type in. The following examples are from notes dictated by the author of this document for his pediatric patients. These errors are actual mistakes made by the transcriptionist which, fortunately, were found and corrected before entered as part of the official, searchable, medical record (Table 5).

What was dictated	What was transcribed
Diet: She eats from all 4 food groups, many fruits, but not many vegetables.	Diet: She eats from all 4 *fruit* groups, many fruits, but not many vegetables.
She had a cord around her neck x 2, although there were no complications at birth…	She had a *quarter on the neck* x2, although there were no complications at birth…
He was given a prescription for an albuterol MDI one to two puffs…	He was given a prescription for an *albumin MDI* one to two puffs…
Her foster mother does braid her hair but reports that it is not very tight.	Her foster mother does *grate her hair* but reports that it is not very tight.
Mom reports that this has been a persistent problem that she has had since birth, and she was much worse yesterday…	Mom reports that this has been a persistent problem *that she has had some spurt,* and she was much worse yesterday…
He has bilateral red reflex. His ears are well situated. He has no nasal flaring. His oropharynx is clear…	He has bilateral red reflex. His ears are well situated. *He has no nasopharynx*. His oropharynx is clear…
She had been on Claritin-D in the past, but is not taking it now because she was told not to take that while she is taking Sudafed.	She had been on Claritin-D in the past, *but is not taking Enalapril* she was told not to take that while she is taking Sudafed.
Her chest is clear to auscultation bilaterally.	Her *check* is clear to auscultation bilaterally.
Her temperature had been taken in the ear.	Her temperature has been taken in the *ER*.
Mom also reports that she is working on potty training him and has a reward system in place…	Mom also reports that she is working on potty training him and *has no work system in place*…

Table 5. What was dictated…what was transcribed. Errors introduced into the medical record can potentially hamper searching as well as cloud meaning.

As humorous as the above examples are, it is unclear how many of these errors go unnoticed and uncorrected. A statistic once mentioned was that physicians, on average, viewed a note for about 1 second before signing it—hardly enough time to accurately review them to make corrections. The examples above were at least ones in which it might be possible to discern the original intended meaning. Unfortunately, some notes are so garbled that the original meaning is lost, as can be seen from an excerpt from the following ER note that did make it into the official medical record (Figure 2).

During her emergency department stay her abdominal pain improved. It was felt most likely she did not have appendicitis and most likely had viral gastroenteritis. Prior to discharge from the emergency department she was offered some Motrin, however *she deferred atrial tachycardia*.

Figure 2. An excerpt from a note taken from a patient's official medical record. Unfortunately, the ending of the last sentence was not corrected and thus the original meaning is lost.

Even though there is no guarantee that EMERSE will find what you want, the frequency of errors is still likely to be low. Additionally, if EMERSE *does* find what you're looking for you can potentially avoid having to read through tens, if not hundreds, of notes to look for something that may or may not exist in a specific patient's record.

THE EMERSE SYSTEM

The following section contains various screen shots of EMERSE with explanations on the features available and how each feature can be utilized.

Figure 3. The EMERSE splash screen.

The first screen you will encounter is the "splash" screen. This screen provides basic version information about EMERSE and it is a good place to check to see when the system was last updated.

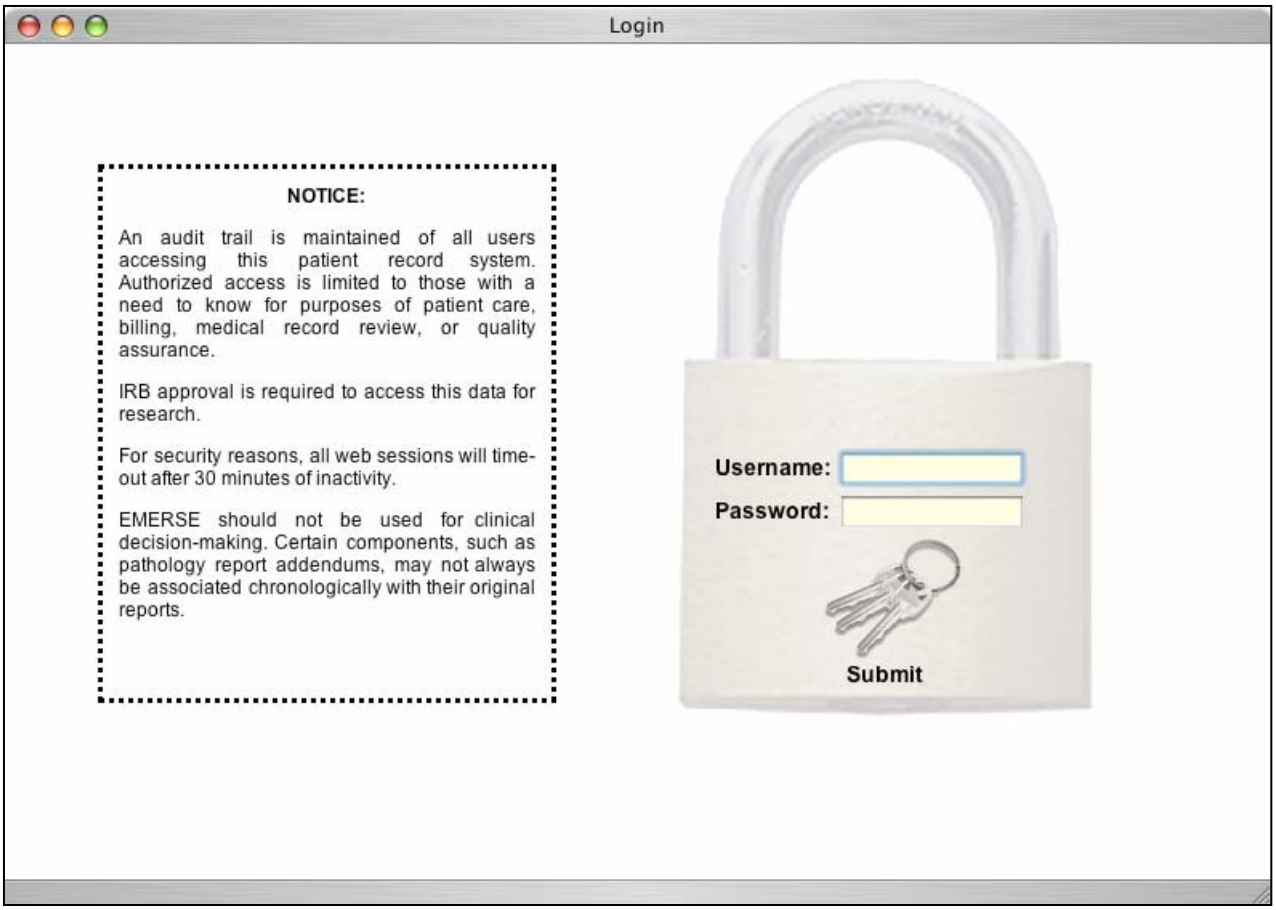

Figure 4. The login screen.

This screen is where you should enter your username and password. Both are case-sensitive. Passwords are stored in a one-way encrypted manner. This means that if you lose your password not even the EMERSE administrator can retrieve it for you, but the administrator can reset it for you.

The "Notice" on the left provides important disclaimer information. The most important points from this notice are:

- An audit trail is maintained
- IRB approval is required for research

This is no different from how CareWeb works, so as a user you should already be familiar with this.

Once logged in, you can change your password by going to the Preferences section.

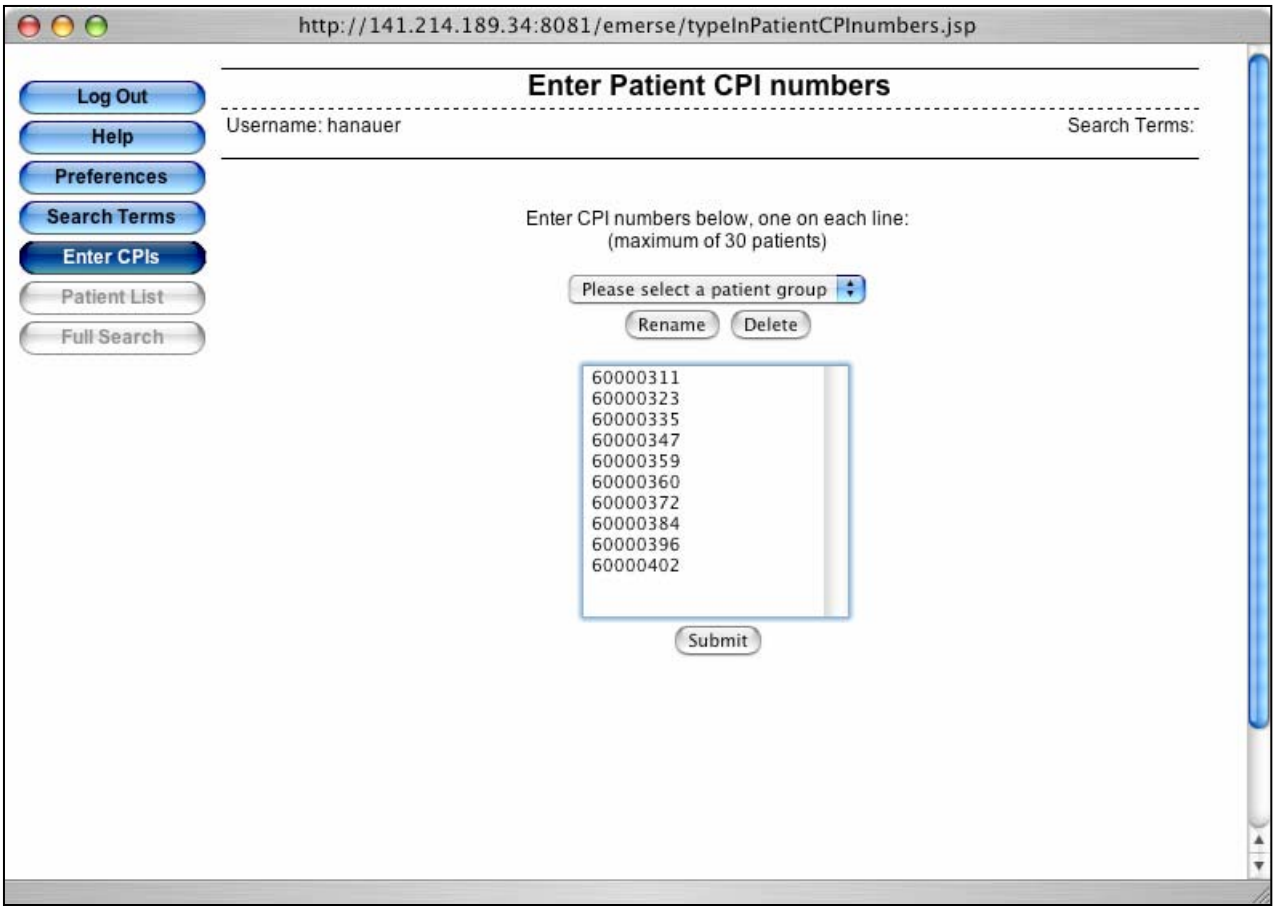

Figure 5. Entering patient CPI numbers.

On this screen you can enter CPI numbers, up to 30 at a time. In the future this upper limit may be increased. You can either type in the numbers one at a time or paste in an entire set of CPI numbers. You can copy and paste such a list from a Microsoft Excel spreadsheet.

The drop-down menu titled “Please select a patient group” exists so that you can save patient lists to recall at a later time. Please see the text associated with Figure 6 on how to save such a list. Once a patient list has been saved you can select it from the drop-down menu and the corresponding CPI numbers will be placed in the box below. Then press submit.

The “Rename” and “Delete” buttons exist to help you manage your saved lists. Simply select the list you are interested in changing and then press either the “Rename” or “Delete” buttons to make the desired changes.

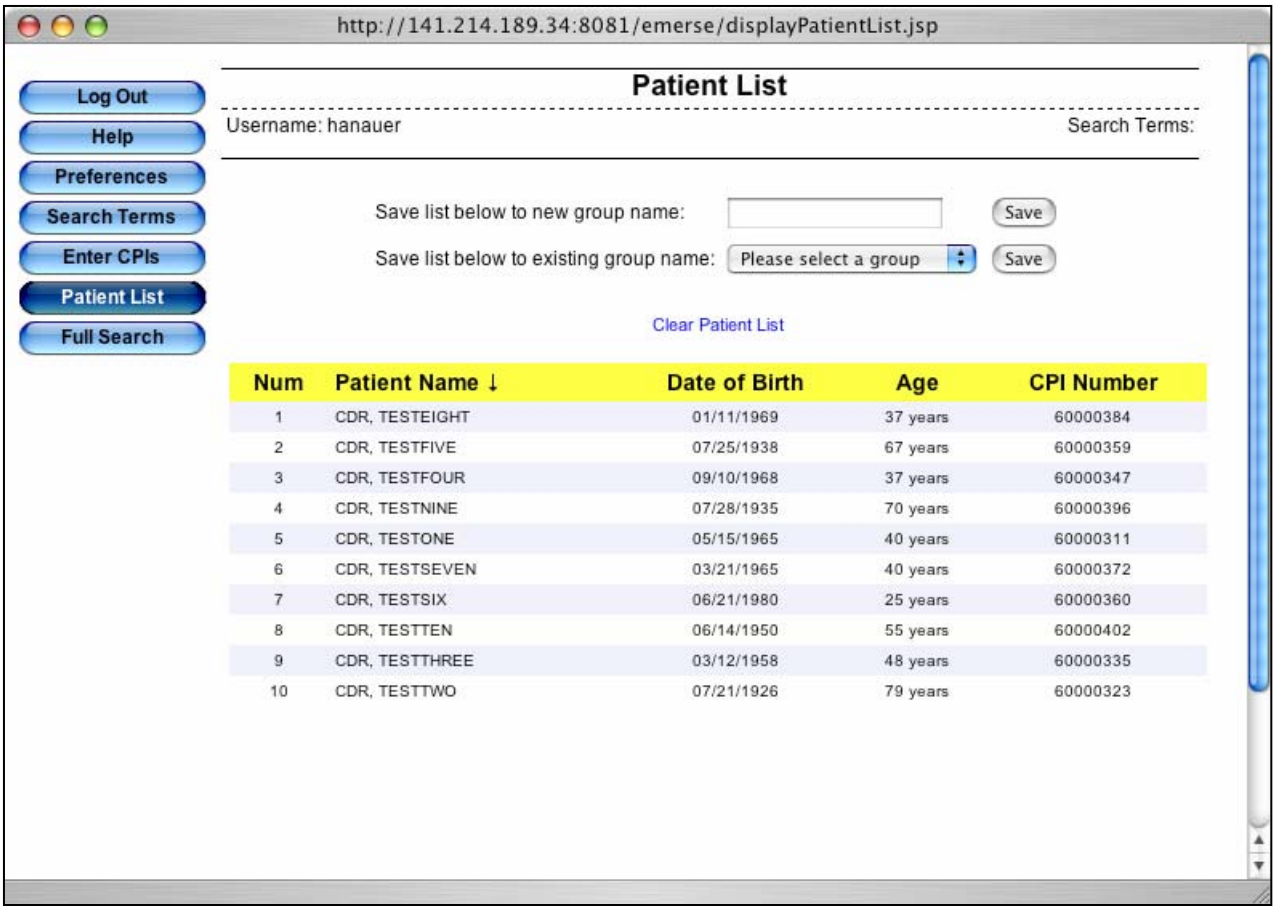

Figure 6. The patient list

This screen represents your patient list after you have entered your CPI numbers. Basic information about each patient is displayed.

Note the small down arrow next to the “Patient Name” column. This signifies that the list is currently being sorted based on Patient Name. To reverse the order in which the list is sorted, click again on the “Patient Name” column. The list will reverse and the arrow will point in the opposite direction. To sort based on a different column, simply click on the column name. The arrow will move next to the name of the column for which the sorting is occurring.

The area above the list should be used to manage your current patient list. If you want to save your current patient list so that it can be recalled at a later time, type a name into the text box and press “Save”. If you want to replace a previously saved list with the current patient list, choose the name of the saved list from the drop-down titled “Please select a group” and press the “Save” button. Please note, however, that this will permanently replace the prior list with the new list.

Clicking on a row will bring up the full demographics for the patient and allows for selecting a search on a specific type of document. This can only be done after a search term has been set.

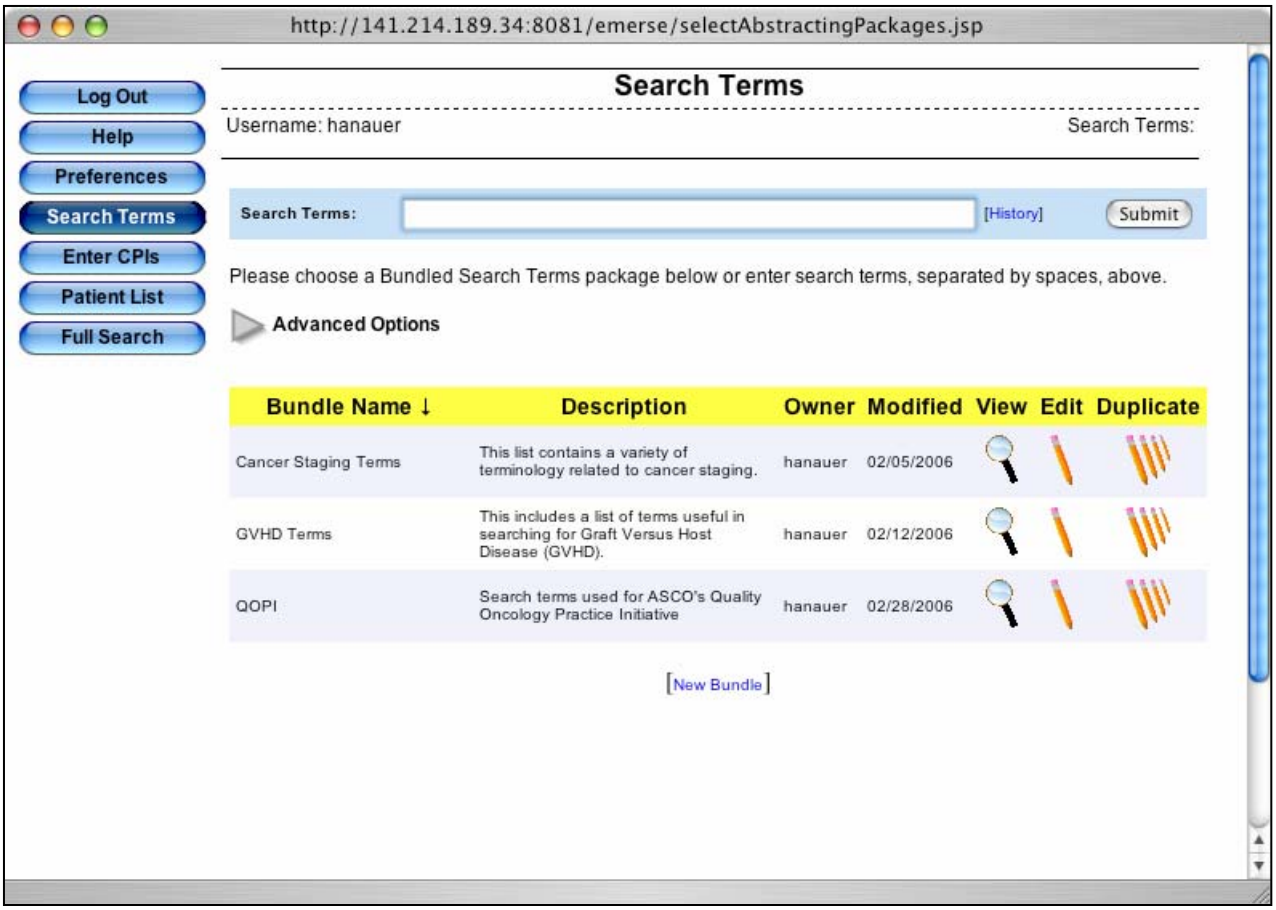

Figure 7. Entering search terms

This screen is where you should enter the terms for which you want to search. Search terms should be entered into the text box; then press the submit button. (Please see the section entitled "Setting Up Search Terms In EMERSE" for details about searching.)

Clicking on the "Advanced Options" arrow will display a list of special modifier characters to aid in your search. (See "Special Characters Used By EMERSE" for more details).

The "History" link will bring up a list of your prior search queries (see Figure 13).

The table at the bottom of the screen allows you to select a search terms "Bundle". This bundle represents a pre-packaged list of terms to perform a search. To choose a bundle, click on its name or description. To view the terms contained in the bundle click on the magnifying glass (see Figure 14). Bundles can also be edited, described on page 26 Click on the "New Bundle" link to create a new, blank bundle to which you can add your own terms.

Figure 8. Search Term Selection

This screen provides confirmation that the search terms were accepted. The upper right corner of the screen shows the search terms entered with their associated highlight colors. If a bundle is selected, the name of the bundle would appear instead.

The entry line for entering a search term is still shown so that you can make any changes or edits. Additionally, if EMERSE detects a spelling error, it will display any suggestions below the entry line under the “Perhaps you meant” heading. If you agree with the suggestions made by EMERSE, simply click on the new suggestion and the corrected version will then be set as your search term. Otherwise, make any edits manually and press the “Submit” button.

The “Alternatives” link is explained later on page 23

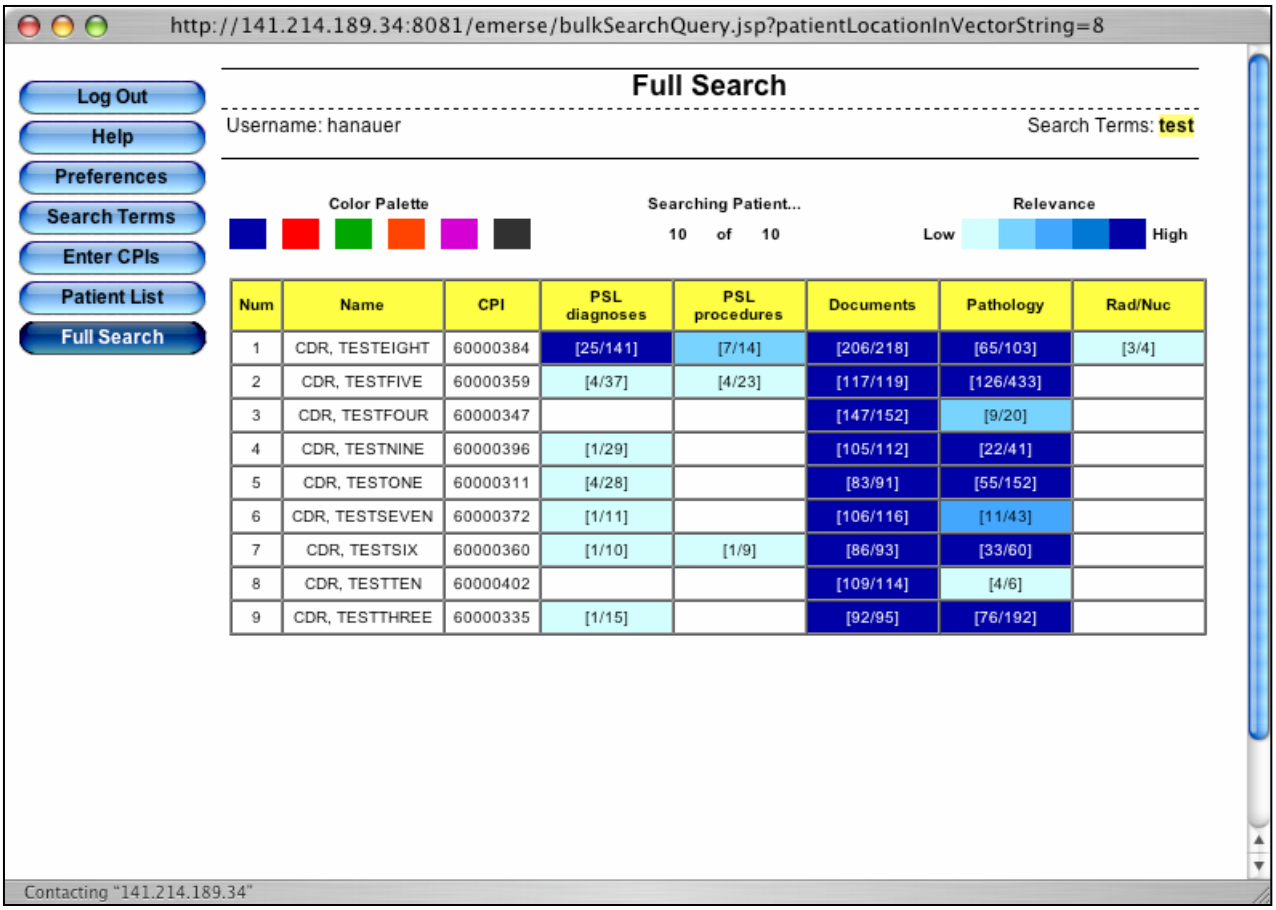

Figure 9. The Full Search window

Once you have set your search criteria, clicking on the "Full Search" button, on the left side of the window will bring up the "Full Search" screen. EMERSE will search each patient, one at a time. While it is conducting the search it will state "Searching Patient…" as can be seen above. Once the "Searching Patient…" text disappears the search is complete.

The table is essentially a "heat map" representing an overview of the search results. Each row represents a patient, and each column represents a type of document. Cells that are blank (i.e. white) have no associated "hits" for that document type and specific patient. Cells with a "hit" are color-coded based on a measure of relevance which is shown by a scale in the upper right above the table. By default, the numbers inside each cell represent the number of documents with a "hit" over the total number of documents for that specific patient. These settings, including how relevance is ranked and the information displayed in each cell, can be changed under the Preferences section. Clicking on a color under the "Color Palette" area simply changes the color scheme used in the heat map.

To drill down and view a list of the hits for a particular patient, click on the cell of interest. You will then see a listing similar to that in Figure 10.

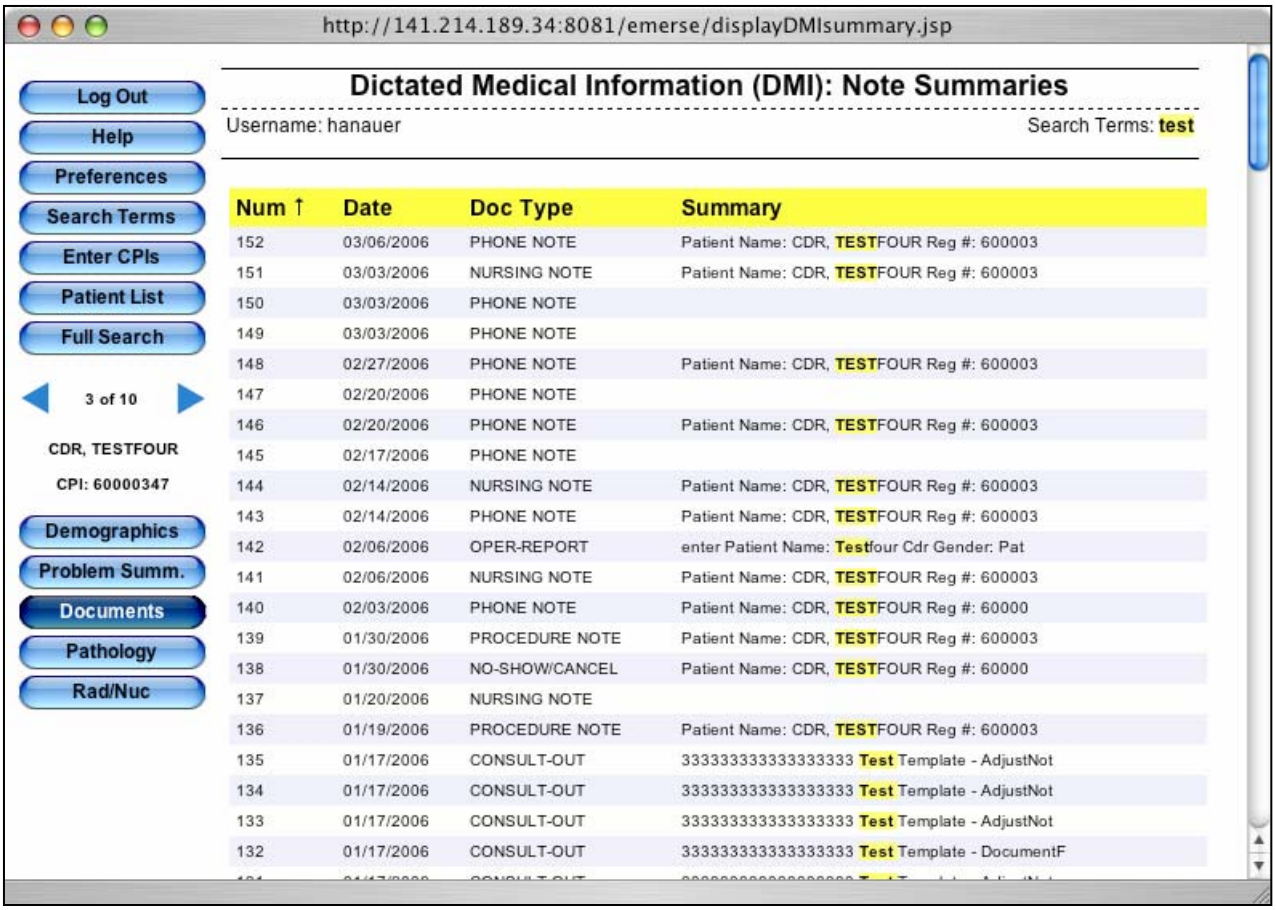

Figure 10. The summary list of documents

This list can be reached either via clicking on the appropriate cell in the heat map (Figure 9) or by clicking on the button on the left of the screen for the type of document of interest for a patient.

By default, documents are listed in reverse chronological order. The arrow next to the column heading represents the column on which the sorting was performed. Similarly to the sorting explained for Figure 6, clicking on a column heading will cause the information to be sorted according to that column.

Each row in the list represents an individual note or document. If the "Summary" section is blank, it means that no "hit" was found for that document based on the search terms entered. Otherwise, if a "hit" was found, the first instance (only a single instance) of a hit is displayed with a small amount of text around it to help establish the context. The length of the surrounding text can be changed in the Preferences.

To view a specific note, simply click on the row of interest.

The arrows underneath the "Full Search" button allow rapid switching between patients for the type of document currently being viewed. The current patient's name and CPI number are displayed beneath the arrows.

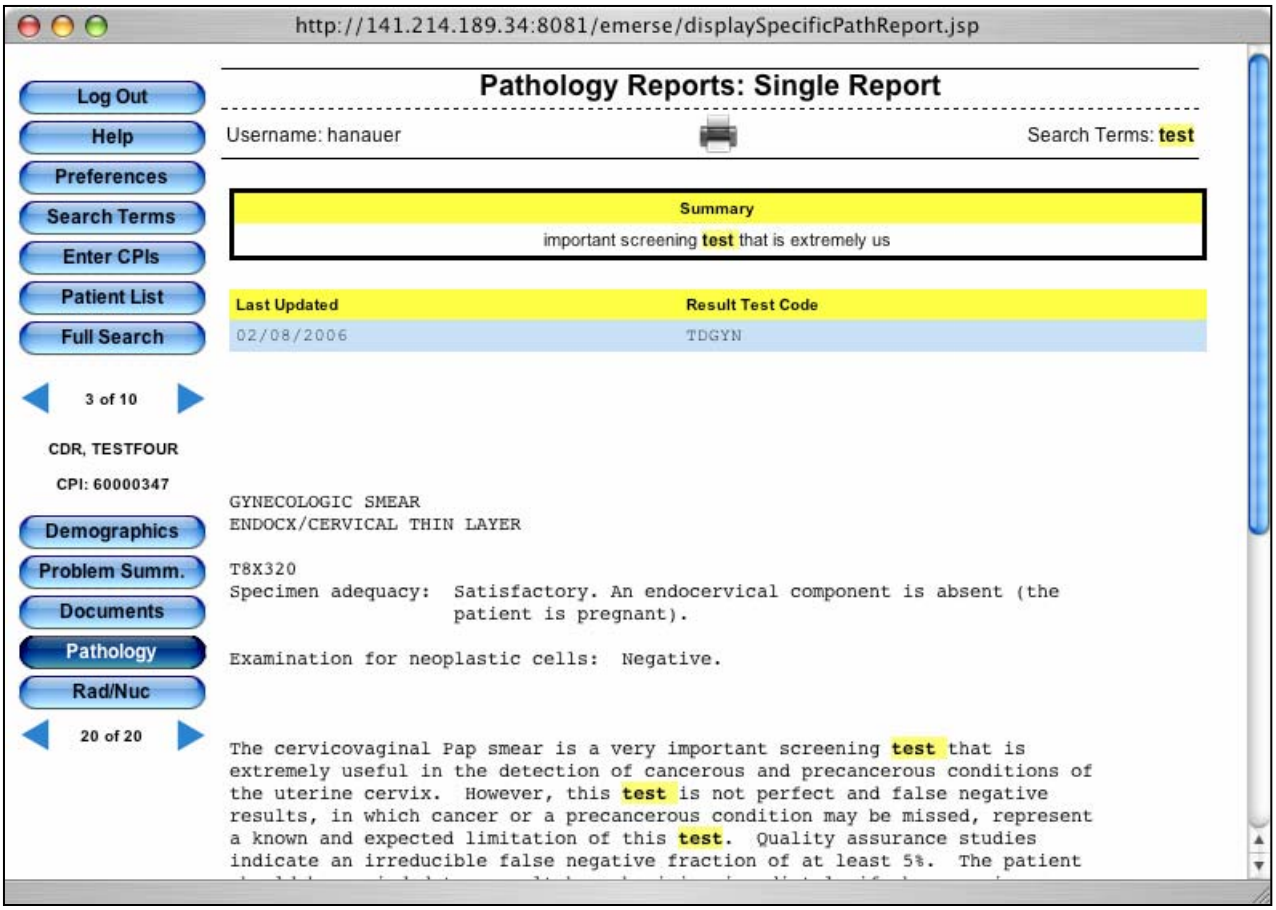

Figure 11. Display of a single document

Single documents are displayed when a row for a document is selected. The note will be displayed and all "hits" in the document will be shown in highlighted colors. Clicking on the printer icon (near the top center of the window) will bring up a new window with the document formatted for printing, with the buttons and menus removed.

The bottom set of arrows (below the "Rad/Nuc" button) allow for rapidly switching between single documents for the selected patient.

The Summary for the single document selected is still shown towards the top of the screen; this provides an area that can be rapidly scanned to determine if any hits exist in the document. If a document is long and the hits only appear at the bottom, looking at the summary will give an indication of whether or not anything might be found by scrolling through the document. If the summary is blank it means that the document contains no hits.

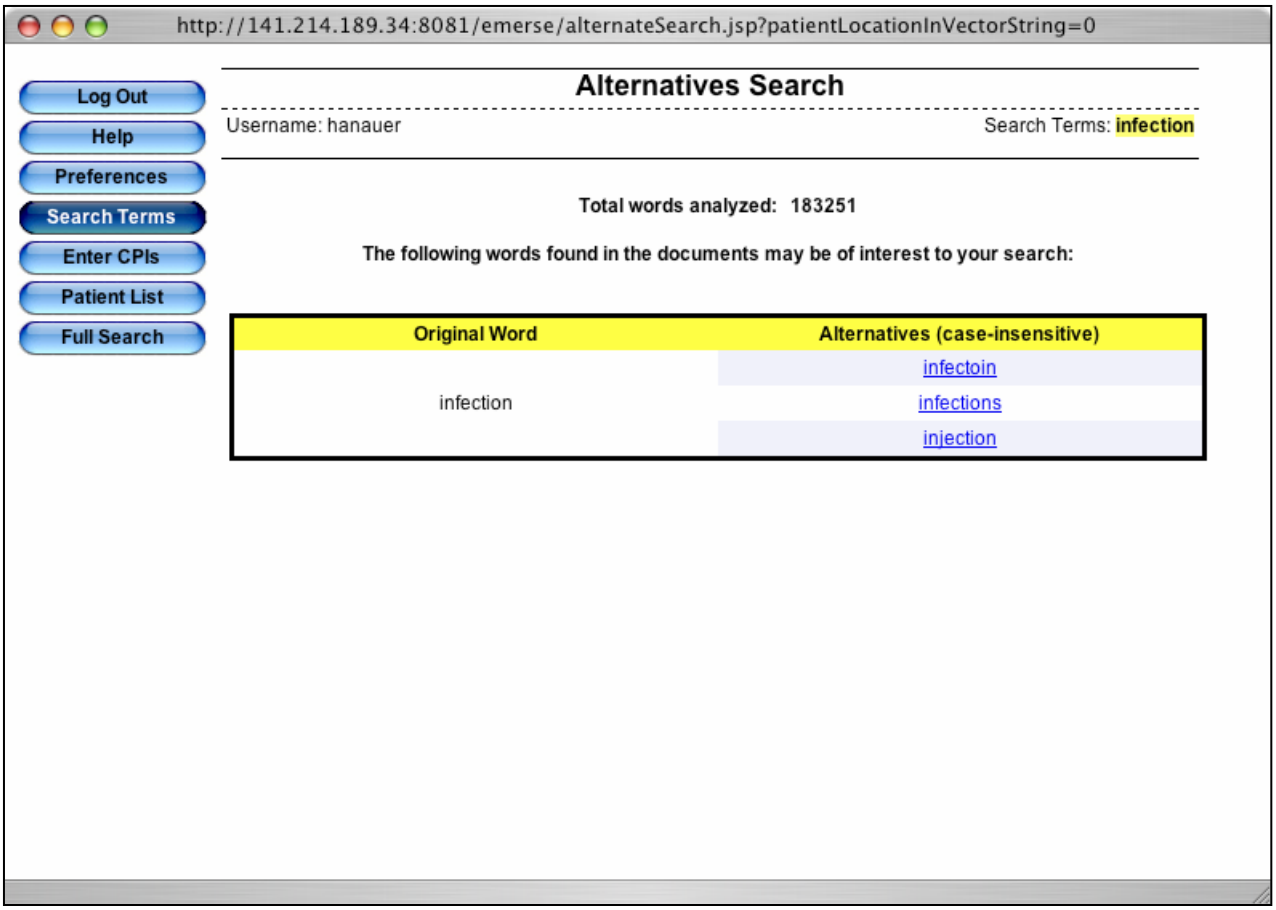

Figure 12. The Alternatives Search

The “Alternatives” feature can be chosen by clicking on the appropriate link (See Figure 8, page 19.

This feature is useful for finding potential spelling errors in the documents so that the search can be refined to include those errors as well. It will catch spelling errors as well as other errors such as “prostrate exam” or “generic screening”. In the case of these two examples, all words are correctly spelled but the words are not the same as those intended for the search.

EMERSE will scan every word in every document for the patients in your current patient list and look for similar words. Any words that are similar will be displayed as can be seen above. In the case of the above example, alternatives for the word “infection” were found. One of these alternatives was the word “infectoin” which represents a spelling error (the others were “infections” and “injection”). To search using an alternative word, simply click on the word; this will then cause it to become the active search term. Then perform a “Full Search” again to locate the notes where the alternative word exists.

By default, EMERSE only shows non-case sensitive alternatives (since the list might be redundant otherwise), but this can be changed in the Preferences section.

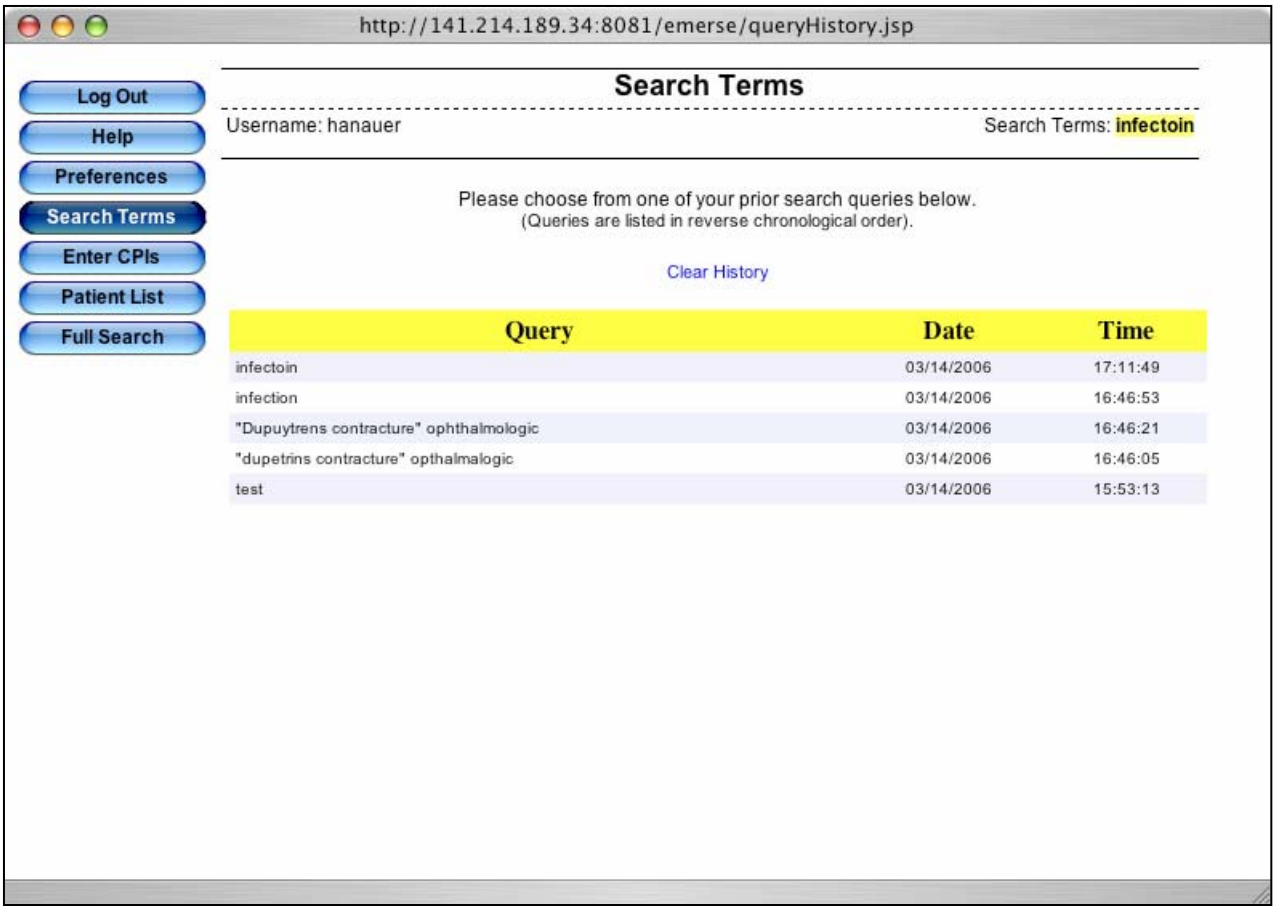

Figure 13. The History window.

The History window can be accessed by clicking on the “History” link (see Figure 13).

This will bring up a list of prior search terms entered. The number of terms saved in the list can be set in the Preferences section. While not shown in the screen shot above, the columns can also be sorted by clicking on the column heading.

To choose a search term from this list, simply click on the associated row. Doing so will set the term in that row as the active search term.

Note that only the search terms typed in will be saved. Selected Search Bundles are not saved in the History list.

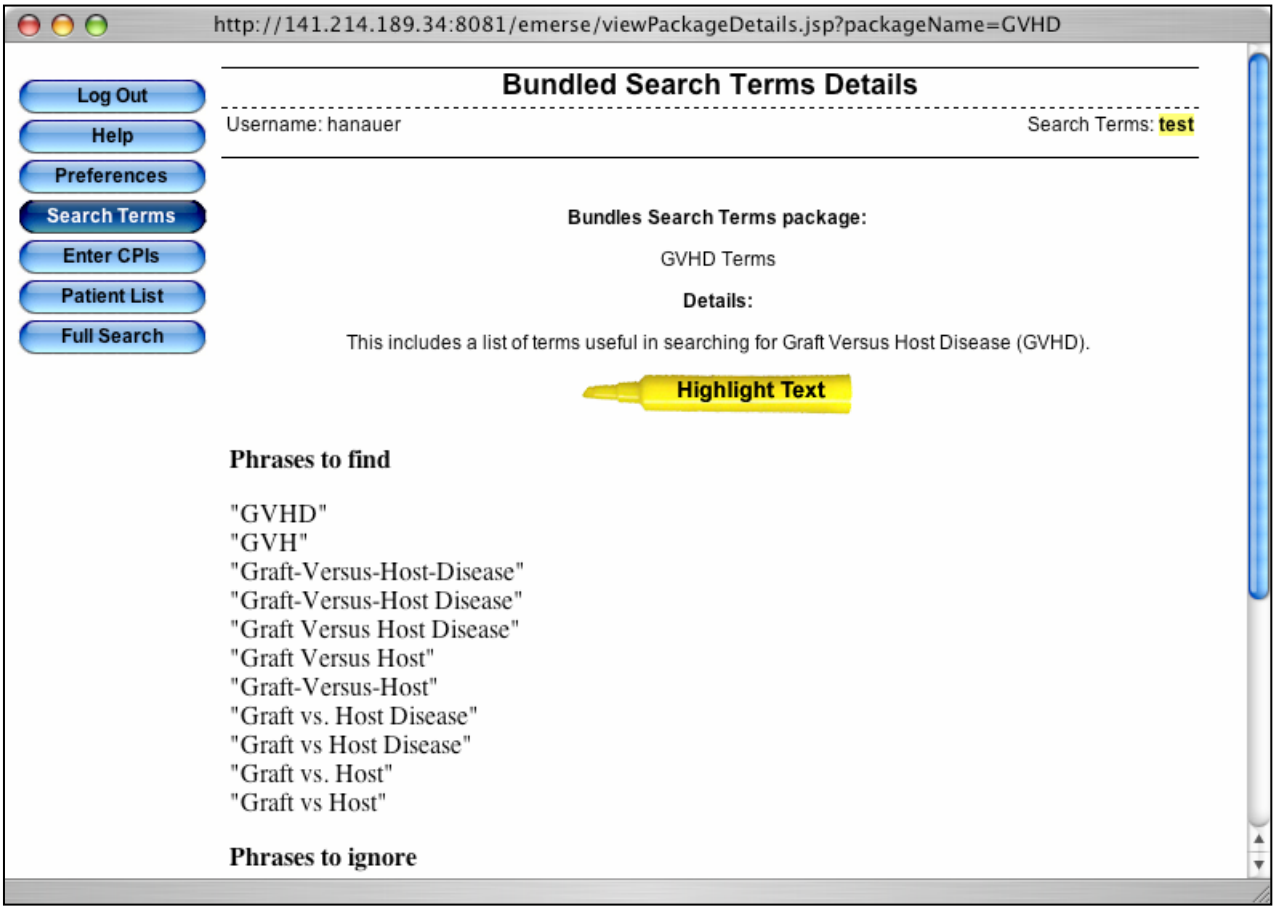

Figure 14. Viewing a bundle.

A Search Terms Bundle can be viewed by clicking on the associated magnifying glass icon as can be seen in Figure 7. This will bring up a list of all of the terms contained in the bundle. The terms are ordered into two lists:

1. Phrases to find: These represent the phrases that EMERSE will search for in the documents and notes
2. Phrases to ignore: These represent the phrases that EMERSE will first ignore before conducting the search (to help eliminate false-positive results)

Clicking on the "Highlight Text" icon will display each search term (each row of the list) in the corresponding color in which it would be highlighted in a document if it were found.

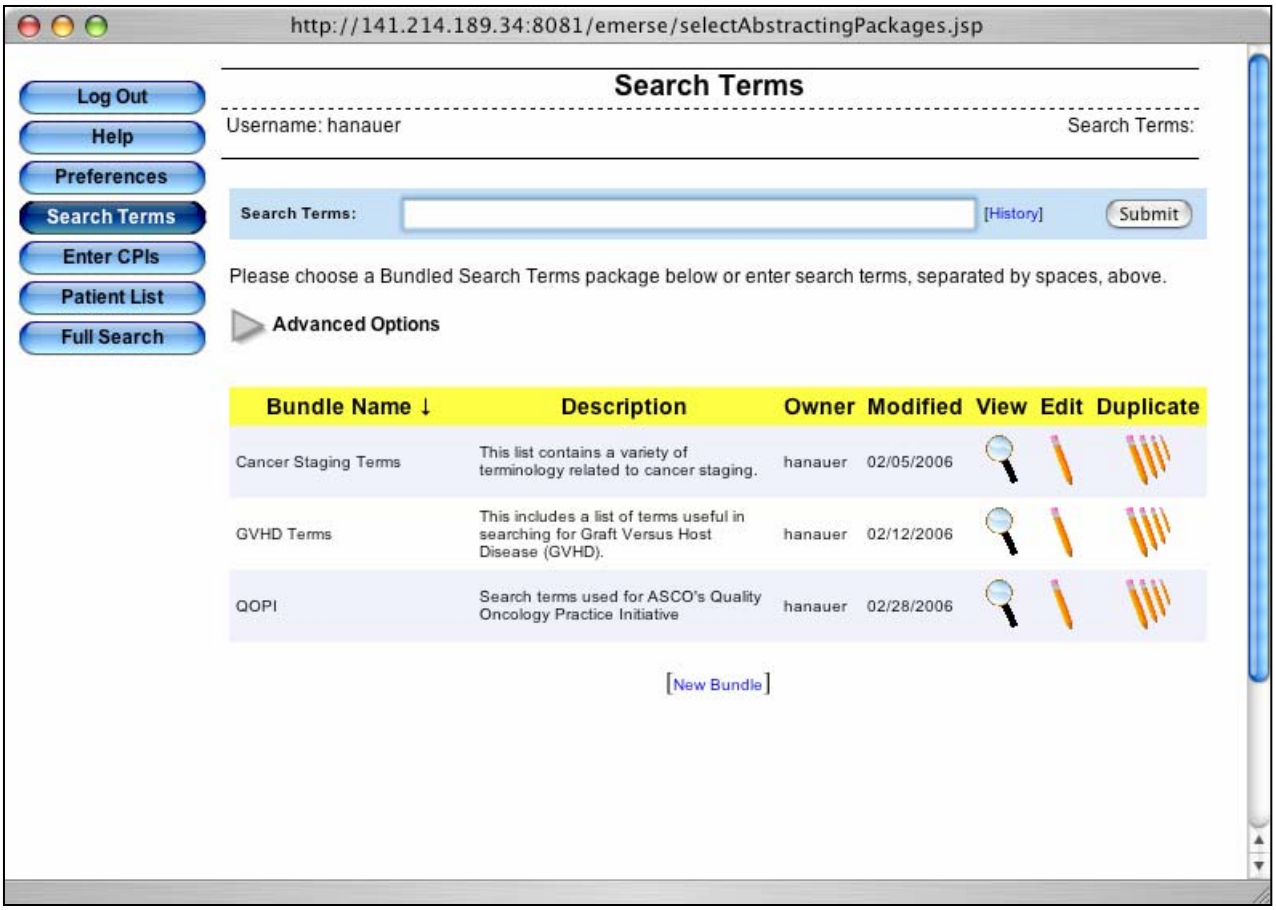

Figure 15. Editing Bundled Search Terms (1)

A powerful feature of EMERSE is the ability to edit or create your own Search Term Bundles. There are several ways in which this can be done.

Only one person at a time can edit a specific Bundle, and that person is designated as the "owner" of the Bundle. If the currently logged in user is also the bundle owner, an icon under the "Edit" column will appear. Clicking on the icon will bring up the editing functions for that bundle.

Another option, if you are not the "owner" of the Bundle, is to click on the "Duplicate" icon associated with the Bundle of interest. This will create your own personal copy of the Bundle that you can edit and change without affecting the original Bundle.

Last, clicking on the "New Bundle" link will create a new, blank Bundle that you can then edit and change as desired.

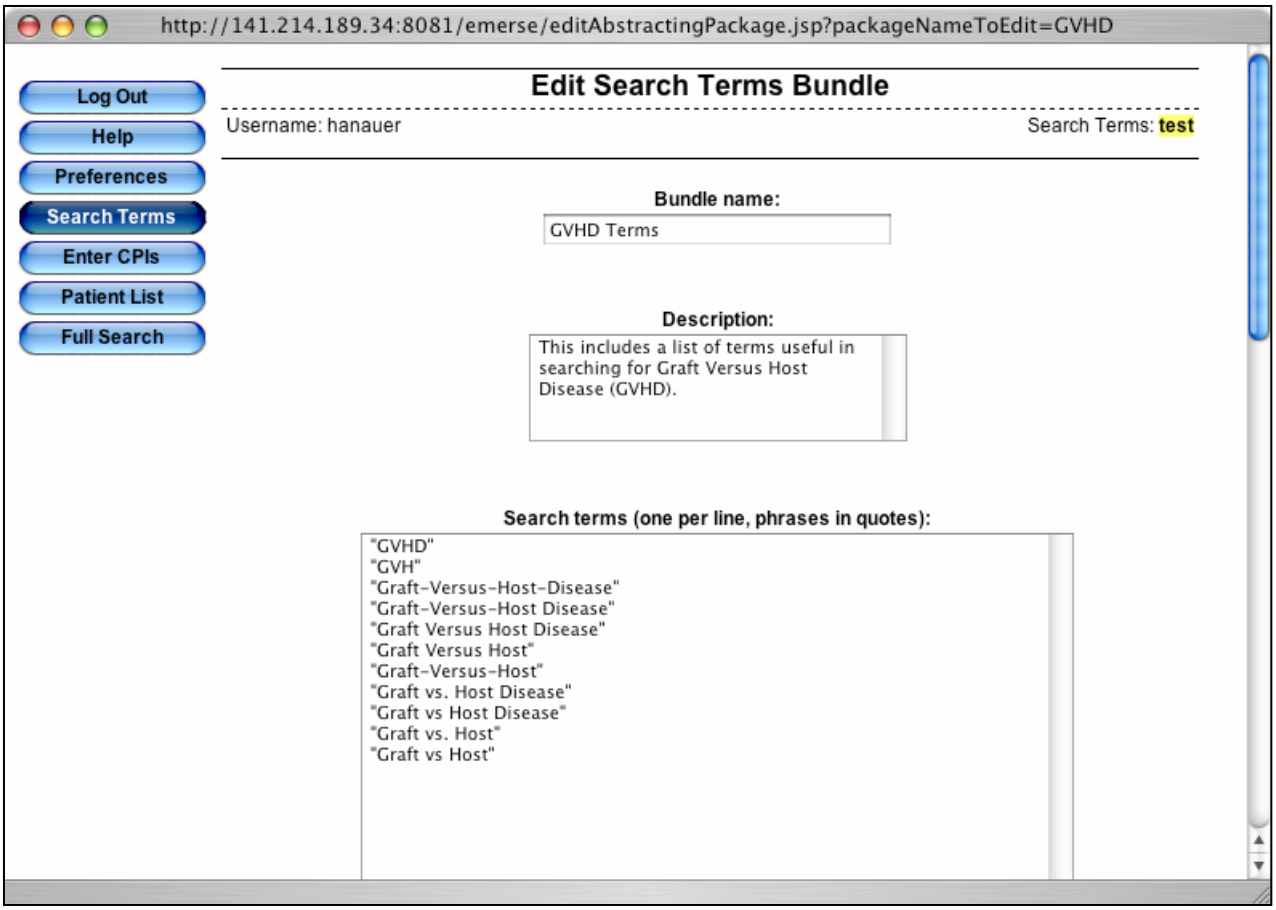

Figure 16. Editing Bundled Search Terms (2)

The edit screen for Bundled Search Terms allows for changing the basic parameters of the Bundle. This includes the

- Bundle Name
- Description
- Search Terms

The Bundle name and description are displayed in the table (Figure 15), so it is best to choose a brief description that will fit well in the space of the table.

There is virtually no limit to the number of search terms you can enter (many thousands of terms could theoretically be entered), but each term or phrase should be on its own line. Additionally, it is important to follow the guidelines for setting up search terms as outlined in the section “Setting Up Search Terms In EMERSE”. Phrases to be ignored can also be added to this list by placing a minus sign in front of the term or phrase.

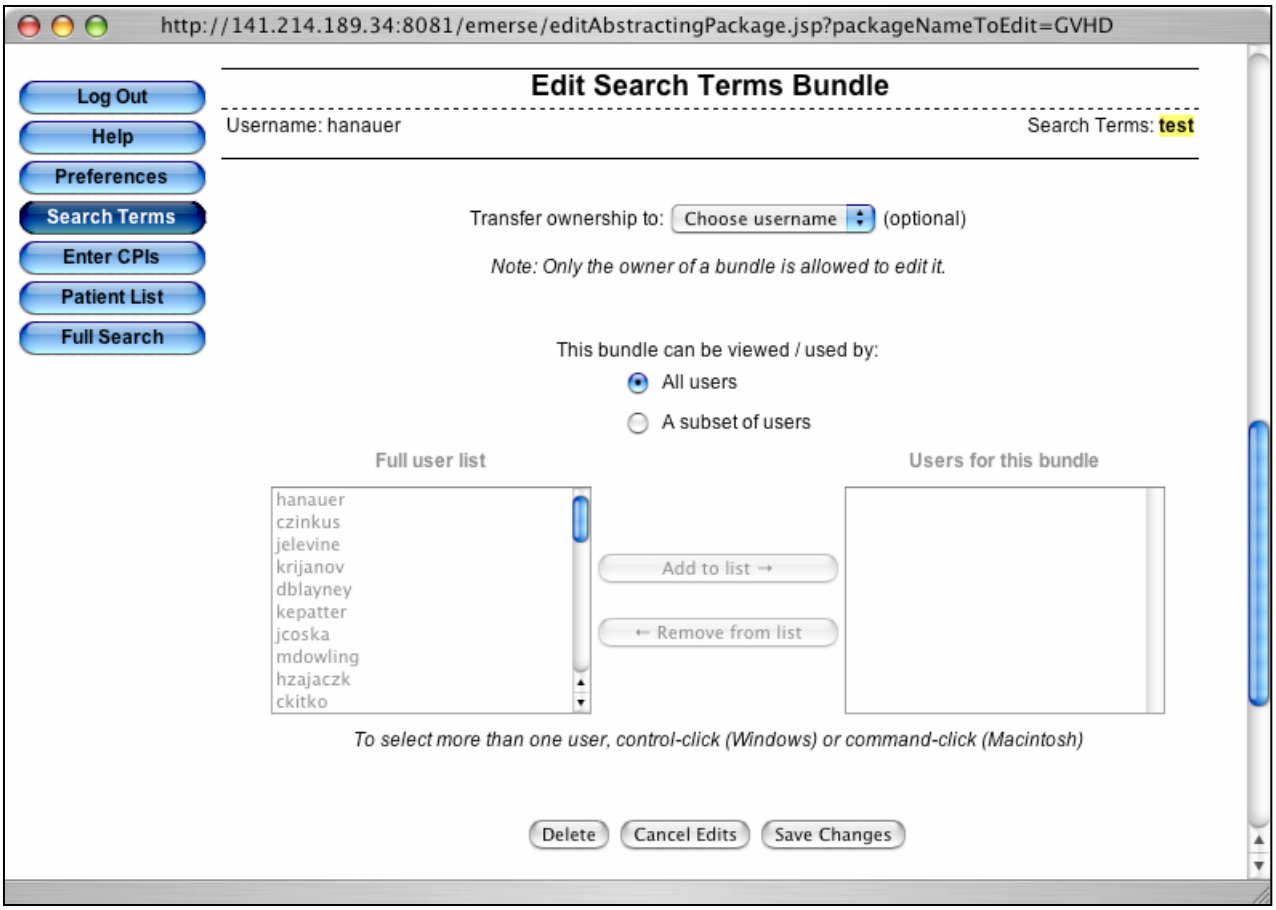

Figure 17. Editing Bundled Search Terms (3)

Scrolling further down the screen from what is shown in Figure 16 reveals a few more options available for a Search Term Bundle.

The option to "Transfer ownership" of the current Bundle is optional. If you are the person editing the Bundle you must, by definition, be the owner of it. If you want to allow another user to edit the same Bundle, you must transfer ownership of the Bundle to that user. Once ownership is transferred only the other user can then edit it and your own editing privileges will no longer exist.

It is also possible to set who is allowed to view the Bundle. These individuals will be able to use and potentially duplicate the Bundle, but they will not be permitted to edit it since they are not the owners. By default, when a new Bundle is created or when a Bundle is copied, only the individual who created or copied the bundle is able to view it. Once a bundle is ready for others to use, you can change it so that all users can view the Bundle or only a subset of users. If a subset of users is selected then you can choose which users can view the Bundle by adding their names from the left-hand box to the right-hand box (see Figure 17, above). If a subset of users is chosen, at a minimum, the owner of the Bundle must be able to view it.

Figure 18. Editing Bundled Search Terms (4)

A final manner in which a search Bundle can be created is by clicking on the “Make Bundle” link, which can be seen to the left of the “Submit” button, above.

The screen shot above can be seen once a search term has initially be submitted.

The utility of making a Bundle from a search term exists when a lot of terms have been entered and it is determined that a Bundle would be a more desirable approach to managing the list of terms or phrases.

Clicking on the “Make Bundle” link automatically creates a new Bundle with the current user as the owner of that Bundle and adds the terms directly into the Bundle. For example, with the terms seen in the screen shot above, the Bundled terms would be listed as:

“heart attack”
“myocardial infarction”
^MI
“chest pain”

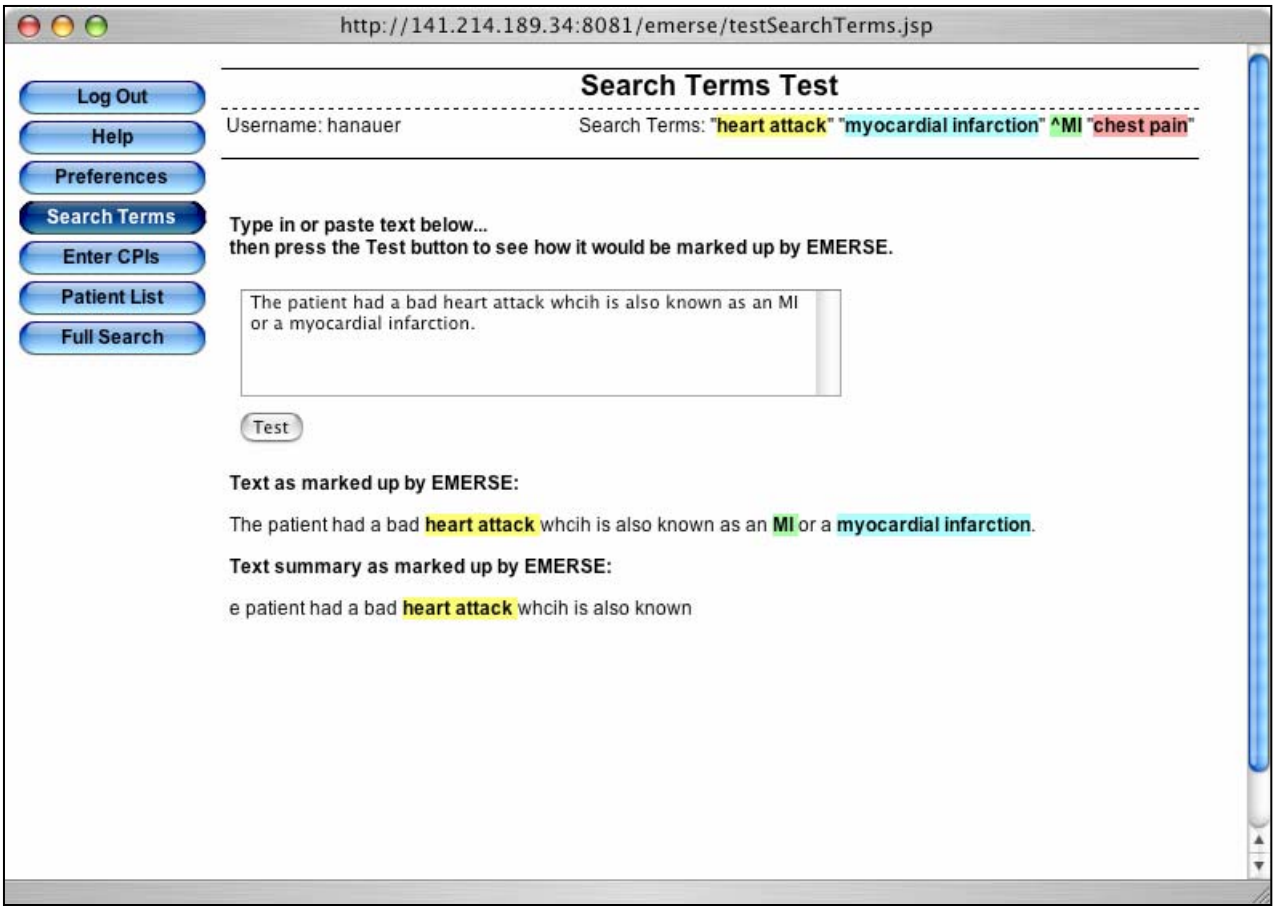

Figure 19. Testing Search Terms

The occasion may arise where a search term, or list of search terms, has been created but the manner in which text will be highlighted is difficult to predict.

A feature exists where it is possible to type or paste text into a text box (as seen above) and then to press the “Test” button where all relevant terms will be highlighted.

To access this feature, click on the phrase “Search Terms:” which is in the upper right corner of the window, as seen above.

The text entered into the text box will show how the text would be marked up by EMERSE and will also show how the text summary would appear (such as when all of the pathology reports are listed for a patient).

SETTING UP SEARCH TERMS IN EMERSE

In order to find what you're looking for, EMERSE offers multiple options for powerful and fast searches. When EMERSE finds a search term it will highlight that term with one of ten different colors to help you find it faster in the document. A description of how to set up your search terms, with examples, follows.

1. The most basic type of search is where you simply enter the terms of interest, separated by spaces.

TERMS: heart pulmonary sinus wrist
The heart was found to be normal and the patient's pulmonary function tests were within normal limits. He also complained of sinus pressure.
Note: By default the spaces between the search terms are considered to mean "or". That is, it will search for "heart" or "pulmonary" or "sinus" or "wrist". In the example above, three of the four terms were found and highlighted. Wrist wasn't found or highlighted, but that didn't affect the ability of EMERSE to find the other words.

2. Sometimes just typing in the words of interest might result in unexpected "false positive" hits.

TERMS: heart attack
The patient was found to be having a panic attack; his heart was found to be normal.
Note: The search criteria were set up to find the word "heart" anywhere or the word "attack" anywhere, which is exactly what EMERSE highlighted. However, this was not likely the desired search result.

3. Putting a phrase in quotes will force EMERSE to search for the words in that specific order (the spaces inside of the quotes are part of the phrase and don't mean "or" like they do when outside of the quotes).

TERMS: "heart attack" "asthma attack"
The patient had a heart attack and not an attack of asthma as he originally thought. When told that it was a heartattack he was surprised. We then explained the signs to look for with a heart-attack.
Note: The phrase "heart attack" was highlighted in two out of three cases. This happened even when the words were joined together without any intervening "white space". The phrase "asthma attack" was not highlighted since that specific wording wasn't used—"attack of asthma" was used instead. The phrase "heart-attack" was not highlighted because of the hyphen between the two words.

4. Think about the concept for which you are searching. There may be synonyms and abbreviations for which you should also search. You should search for all possibilities or synonyms.

TERMS: "world health organization" WHO
This patient meets criteria outlined by the World Health Organization. The patient, who is 70 years old, meets the WHO definition of kwashiorkor.
Note: By default, all searches are case-insensitive.

5. To force EMERSE to search for a term in a case-sensitive manner, add a carat (^) symbol directly in front of the word, without intervening spaces.

TERMS: "world health organization" ^WHO
This patient meets criteria outlined by the World Health Organization. The patient, who is 70 years old, meets the WHO definition of kwashiorkor.
Note: Now only the desired "WHO" term was highlighted. Carat symbols can also be added in front of specific words in quoted phrases as well.

6A. EMERSE, by default, will search on any subset of a word of phrase that matches. Be careful about your choice of search terms in order to be inclusive.

TERMS: "pulmonary function tests"
The patient had a normal pulmonary function test.
Note: "pulmonary function tests" is not a subset of "pulmonary function test" because of the extra "s". As a result, nothing was found in this case.

6B. By using the singular version of the desired phrase, there is less likelihood that something will be missed.

TERMS: "pulmonary function test"
The patient's current pulmonary function test was abnormal, which is different from his previous pulmonary function tests. Further pulmonary function testing is advised.
Note: In this case EMERSE finds all terms with the phrase "pulmonary function test" regardless of the ending of the word "test".

7. Because EMERSE searches any subset of text by default it could result in false positives.

TERMS: ^ALL
DIAGNOSIS: ALL (acute lymphoblastic leukemia) ALLERGIES: Allergic to all types of dogs.
Note: EMERSE did not find the lower case "all" but it did incorrectly highlight the capitalized "ALL" in allergies. There are several potential ways to avoid this problem.

8. You can instruct EMERSE to ignore certain words before conducting the search by placing a minus sign (-) in front of the term or quoted phrase. This can help avoid false positives.

TERMS: -allergies ^ALL
DIAGNOSIS: ALL (acute lymphoblastic leukemia) ALLERGIES: Allergic to all types of dogs.
Note: Now only the proper "ALL" was found.

9. You can force EMERSE to search for a whole word as opposed to only part of a word by adding a tilde (~) character in front of it. This means that other letters can't exist immediately on either side of the word for which you search.

TERMS: ~^ALL
DIAGNOSIS: ALL (acute lymphoblastic leukemia) ALLERGIES: Allergic to all types of dogs.
Note: This method may also work for eliminating some false positive hits.

10. Again, be careful about how you set up your search terms.

TERMS: head CT
Clinical history: Head trauma secondary to impact with baseball bat. DIAGNOSIS: SKULL FRACTURE
Note: "CT" was not set to be case sensitive in this search which is why EMERSE picked up the "ct" in "impact". Even specifying capitalized "CT" however would still have picked up the "CT" in "FRACTURE". Thus, using the search term "~^CT" might be best in this case.

11A. Sometimes you may only want to find terms if they all occur together in the document, but not necessarily in a particular order. Adding a plus (+) sign in fron of those words forces EMERSE to find them only if they occur together in the document.

TERMS: "renal artery stenosis" +renal +artery +stenosis vomiting
History: Patient has had forceful vomiting for 12 hours. Diagnosis: pyloric stenosis
Note: The word "stenosis" wasn't highlighted because it had to appear in the same document where the words "renal" and "artery" also appeared. However, "vomting" was highlighted because a plus sign (+) was not placed in front of it as a search term.

11B. An advantage of using the plus signs (+) is that you can find phrases/ concepts even if the word order is variable.

TERMS: "renal artery stenosis" +renal +artery +stenosis vomiting
Significant vomiting has occurred in this patient, although it is unlikely due to stenosis of his renal artery.
Note: The exact phrase "renal artery stenosis" was not found, but by adding a plus sign in front of the words separately EMERSE was still able to find the desired concept

12A. Sometimes you may not be sure of what you are looking for or how something will be worded. In that case you can use a wildcard symbol—the asterisk (*). Using the * symbol means that anything can match it.

TERMS: No evidence *.
No evidence of carcinoma. No findings suggestive of cancer.
Note: The use of the wildcard above means that it will match anything with the words "No evidence" on one side and a period (.) on the other side.

12B. When EMERSE tries to pattern match, it performs the match from left to right. Once it finds something on the left, it will continue to match everything in between as long as there is still something that will match on the right.

TERMS: car*oma
He learned about his diagnosis of carcinoma while driving his car from Michigan to Oklahoma.
Note: Using a wildcard usually won't result in such bizarre highlighting, but this example demonstrates a potential downside to matching anything.

13A. You can limit the range of what the wildcard can match by using a different wildcard symbol: @. The "at sign" (@) will only match part of the single word in which it resides.

TERMS: car@oma
He learned about his diagnosis of carcinoma while driving his car from Michigan to Oklahoma.
Note: Now only the proper term was highlighted.

13B. More examples using the @ symbol.

TERMS: in@on hear@
Upon further inspection of her infection, it was clear that her heart was not affected but she didn't hear us because she was shearing her sheep which affected her hearing.
Note: By using the @ wildcard, each match is limited to only a single word. We could avoid having the word "shearing" partially highlighted by using the search term "~hear@" instead.

14. If you choose to perform a search that uses one of the special symbols ^, *, or @, you must let EMERSE know this by putting a backslash (\) in front of the symbol.

TERMS: per mm\^3
3500 cells per mm^3
Note: If you wish to use the other symbols (~, +, -) simply place them inside of a quoted phrase.

15A. You can use the ^, *, or @ inside of a quoted phrase. The ^ symbol specifies case-sensitivity for only the word it is placed in front of.

TERMS: "he * have ^ALL"
He is presumed to have ALL even though he does not have all of his test in yet.
Note: Adding the ^ symbol in front of the "ALL" inside the quoted phrase allowed EMERSE to pick up the first "he … have ALL" phrase but not the second "he … have all" phrase.

15B. When EMERSE matches using the * wildcard, it starts from the left and works its way to the right. This can sometimes result in text getting highlighted in an unexpected manner, as is the case when the sentence clauses above (15A) are reversed.

TERMS: "he * have ^ALL"
He does not have all of his tests in yet even though he is presumed to have ALL.
Note: Because of the way the search was set up, EMERSE begins by matching the first "he" and continues to match everything in-between until it comes across the capitalized "ALL".

16. As you have observed in the above examples, you can combine various symbols to specify your search criteria.

TERMS: -~^A@e a@e
Apple apple Advantage advantage Archipelago archipelago
Note: The first search term above commands EMERSE to ignore all words that begin with a capitalized "A" and end with a lower case "e". The second search term commands EMERSE to then search for any subset of a word (since the ~ symbol is not in front) that has an "a" followed by any other text and then an "e".

17. The final symbol that can be used with EMERSE is the dollar sign ($). Use this symbol if you want to use a regular expression. If you don't know what a regular expression is you can probably ignore this feature. Regular expressions are powerful, computer-science based methods to search through text. Their downside is that they can be complicated to set up. If you want to use a regular expression, simply use $"regex" where regex is the regular expression. All characters are allowed (and don't need to be escaped) except for the double quotation mark, which is not allowed.

TERMS: $"\b(?:((?i)heart)\W+(?:\w+\W+){0,5}((?i)attack)|((?i)attack)\W+(?:\w+\W+){0,5}((?i)heart))\b"
Heartattack Heart attack Heart one two three four attack Heart one two three four five attack Heart one two three four five six attack
Note: The regular expression above allows for a case-insensitive search of the words "heart" and "attack" and specifies that they can be separated by as few as 0 words and at most 5 words.

SPECIAL CHARACTERS USED BY EMERSE

The following table outlines the special characters recognized by EMERSE to specify how a search query should be conducted, what those symbols mean and where they can be used:

Symbol	Meaning	Where To Use
^	Case-sensitive	In front of any word, even words inside quoted phrases
+	All words with a plus sign must be found together for it to be a 'hit'	In front of an individual word or in front of a quoted phrase
-	Ignore word or phrase	In front of an individual word or in front of a quoted phrase
~	Treat term as a distinct word	In front of an individual word or in front of a quoted phrase
*	Wildcard, matches entire phrase	Inside or edge of individual word or inside a quoted phrase
@	Wildcard, matches single word	Inside or edge of individual word or inside a quoted phrase
$	Regular expression	In front of a quoted regular expression

Table 6. The special symbols (or flags) recognized by EMERSE to specify how a search should be conducted.

IMPROPER USE OF SPECIAL CHARACTERS

The following table list symbols that cannot be combined or locations in which the symbols will not have their intended meaning:

Not allowed	Reason
`+-word`	The plus and minus symbols cannot be combined. You can only find words together with the plus symbol. You cannot ignore words together with the plus symbol.
`+-"quoted phrase"`	The plus and minus symbols cannot be combined. You can only find phrases together with the plus symbol. You cannot ignore phrases together with the plus symbol.
`^"quoted phrase"`	Words must be made case-sensitive individually. An entire phrase cannot be made case-sensitive in this manner.
`*"quoted phrase"`	Wildcards must be part of a word or quoted phrase and cannot modify the entire phrase itself.
`@"quoted phrase"`	Wildcards must be part of a word or quoted phrase and cannot modify the entire phrase itself.
`~$"regular expression"`	You can specify that the regular expression should deal with separate words in the regular expression itself.
`^$"regular expression"`	You can specify case-sensitivity in the regular expression itself.
`*$"regular expression"`	You can specify wildcards in the regular expression itself.
`@$"regular expression"`	You can specify wildcards in the regular expression itself.

Table 7. Combinations or location of special symbols that are not permitted in the EMERSE system. Additionally tabs are not permitted, although they can still be specified, if needed, through the use of a regular expression.

ADDITIONAL SEARCHING TIPS

1A. The order in which you list your search term matters if there is potential overlap. You should list the most specific search term first, followed by less specific terms because EMERSE starts to match up terms with the left-most term and works it way through from left to right.

TERMS: "breast carcinoma" "breast ca" carcinoma ca
This patient has evidence for breast carcinoma. Please evaluate for breast ca.
Note: EMERSE takes the list of search term in order from left to right. When it finds a match EMERSE temporarily sets the matching aside and then looks for any other matches for the remaining text. This can be problematic if search terms are not in a compatible order, as seen in the next example.

1B. When search terms are not put in a compatible order, the text might not be highlighted properly.

TERMS: carcinoma ca "breast carcinoma" "breast ca"
This patient has evidence for breast carcinoma. Please evaluate for breast ca.
Note: In general it is unlikely that EMERSE would miss finding something important if the search terms are out of order, but the highlighted text might not be as specific as it could be. In this example EMERSE first looked for the word "carcinoma" and when it found it EMERSE highlighted the word and set it aside. It then looked for the word "ca" and found it at the end of the sentence and set that aside. When EMERSE then searched for "breast carcinoma" it was unable to find it since the original "carcinoma" term had already been set aside. The same is true for the "breast ca" term.

2. All negative search terms to be ignored (those with a "-" in front) are ignored before the other search terms are found regardless of where they are in the list; however, the order among the negative search terms themselves still does matter for the same reasons as normal search terms.

TERMS: "chest pain" -attack pain -"heart attack" heart
`Patient with chest pain caused by a heart attack`
Note: EMERSE first ignored the word" attack" and set it aside. Then when it tried to ignore the phrase "heart attack" it was unable to find it since the word "attack" had already been ignored and set aside.

3. Searches for terms should be conducted using a "cascading" method in order to ensure that the desired term(s) are found. To do this, search terms should be listed from most specific to least specific. When EMERSE is unable to find the first term it will "cascade" to the next term to try to find a match. It will continue this process until the next best phrase matches or no matches can be found.

TERMS: "carcinoma of the breast" "breast carcinoma" "breast cancer" "breast ca" "BR CA"
`65 yo F with BR CA.`
Note: The search was conducted in the following cascading manner. *EMERSE will search for…* *"carcinoma of the breast" …if that fails to match it will then try…* *"breast carcinoma" …if that fails to match it will then try…* *"breast cancer" …if that fails to match it will then try…* *"breast ca" …if that fails to match it will then try…* *"BR CA" …if that fails to match it will give up the search*

4. The cascading method could also be useful if you are concerned about spelling errors.

TERMS: complications complication complicat comp
No complicatoins were noted, but it was not a comprehensive review.
Note: This method helped pick up the misspelled word "complicatoins" (there was a transposition of the "i" and "o") but it also resulted in a false-positive hit for the word "comprehensive" which is a potential downside to this technique.

FURTHER INFORMATION
(How to get help)

For questions, concerns, comments, praise, or complaints, please contact David Hanauer.

E-mail: *hanauer@umich.edu*

Phone: 734-615-0599

NOTES