MEDINFO 2004
M. Fieschi et al. (Eds)
Amsterdam: IOS Press

Patterns of Usage for a Web-Based Clinical Information System

Elizabeth S. Chen, James J. Cimino

Department of Biomedical Informatics, Columbia University, New York, NY USA

Abstract

Understanding how clinicians are using clinical information systems to assist with their everyday tasks is valuable to the system design and development process. Developers of such systems are interested in monitoring usage in order to make enhancements. System log files are rich resources for gaining knowledge about how the system is being used. We have analyzed the log files of our Web-based clinical information system (WebCIS) to obtain various usage statistics including which WebCIS functions are frequently being used. We have also identified usage patterns, which convey how the user is traversing the system. We present our method and these results as well as describe how the results can be used to customize menus, shortcut lists, and patient reports in WebCIS and similar systems.

Keywords:

Needs Assessment, Artificial Intelligence, Information Systems

Introduction

In order to develop appropriate and effective clinical information systems, researchers and developers must have insight on what information is needed and used by clinicians when performing their clinical tasks. Studies of clinician information needs have typically used qualitative methods such as observations and surveys to gain this knowledge [1].

Log file analysis is a quantitative method that can be used to learn what and how information is being accessed in a system. We have performed an analysis of clinical information system (CIS) log files to gain an understanding of users' needs and their usage patterns. The results of this CIS log analysis can be used in a number of ways. We describe our methods, present a subset of results obtained, and discuss how these results can be used for enhancing our Web-based clinical information system as well as for guiding the development of similar systems.

WebCIS

WebCIS is the Web-based clinical information system at New York Presbyterian Hospital (NYPH) that provides clinicians with access to patient records [2,3]. It also contains links to sources of health knowledge such as Micromedex and PubMed. WebCIS displays information from ancillary, registration and ambulatory systems. Each page of WebCIS consists of several frames (Figure 1). The topmost frame displays demographic information for the selected patient as well as links for accessing other patients' records. The leftmost frame contains a menu of functions to access patient data. These include links to different department results such as laboratory, radiology and cardiology. The central frames contain the data requested by the user based on the links selected. Additionally, pop-up windows are used for certain functions such as graphs and spreadsheets.

As part of the NYPH security management scheme, log files are maintained to keep track of WebCIS user activity. An audit system has been developed to monitor these logs to determine if inappropriate or unauthorized activity is occurring as well as to monitor the basic operational aspects of WebCIS [4,5].

System Usage

Past behavior can be used to predict future behavior. In every domain, systems are commonly monitored to observe usage. Understanding how systems are currently being used is essential to making appropriate enhancements. Log file analysis can be used to gain an understanding about all system users [6]. Systems typically maintain log files to keep track of every user's actions. These files can be analyzed to obtain statistics about system users and their usage patterns.

In the health care setting, log file analysis has been used to monitor the use of a variety of Web-based systems including biomedical digital libraries [7,8,9,10]. The Web server log files of these systems are analyzed to obtain statistics such as the number of hits, single IP addresses in a time period, most requested pages, and errors. These results provide ideas on how users search the library and what information they are using. This knowledge can then be used to determine what information to include and how to organize it within the system to better meet the users' needs.

Methods

The analysis of WebCIS log files uses clinical information system (CIS) mining, a technique we have developed that is based on data mining and Web usage mining [11]. This technique consists of four phases: data selection, preprocessing, pattern discovery, and pattern analysis.

In the data selection phase, we obtained WebCIS log files spanning the period January 1, 2002 to December 31, 2002. Each log record contains 7 fields: timestamp, application name, user ID, IP (Internet Protocol) address, MRN (Medical Record Number), data type, and action.

Preprocessing tasks involved de-identifying, cleaning, enriching, and transforming the data in the WebCIS log files. One transformation of the data was to sort log records into user sessions (Figure 2). These sessions depict how the user moves from one piece of information to another in the patient record from login to logout. Each data type in the session represents the WebCIS function or patient information being accessed. Because these data types are at a fine level of granularity, we further transformed them to an intermediate level by replacing each reference to a patient-specific result with ".KEY.". By representing the data at this coarser level of granularity, generalizations can be made about user behavior.

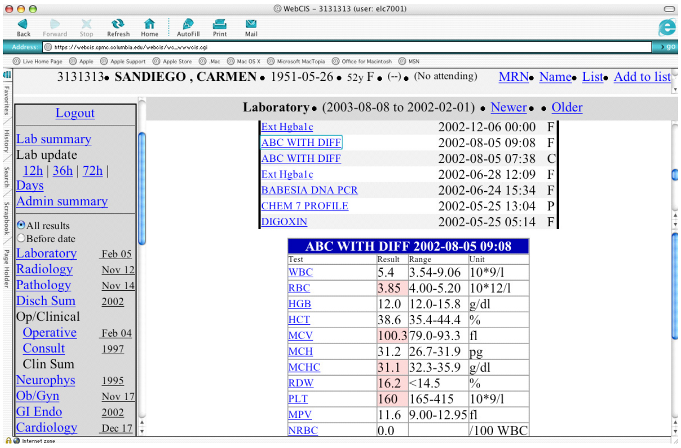

Figure 1 - WebCIS screenshot. The user selected «laboratory» and then chose to view «ABC WITH DIFF»

Pattern discovery tasks included:

- Descriptive statistical analysis: obtain basic statistics about users
- Sequential pattern analysis: identify patterns of data types that occur sequentially in user sessions
- Association rule generation: identify patterns of data types that occur together in user sessions
- Path analysis: visualize how users traverse WebCIS
- Classification: obtain statistics and patterns that are specific to a class of users

With pattern analysis, the results from each of the pattern discovery tasks were aggregated, summarized and reported.

Results

We report results from two of the pattern discovery tasks: the descriptive statistical analysis and sequential pattern analysis.

Descriptive Statistics

From one year's worth of WebCIS log files, we obtained statistics for over 7,000 users viewing more than 400,000 unique patient records. We analyzed about 3,000,000 user sessions containing more than 35,000,000 data types. Table 1 contains a list of the more frequent data types.

Sequential Patterns

With sequential pattern analysis, we analyzed the intermediate-level user sessions to identify sequential patterns of varying length (n-sequences). "n" represents the number of data types found in the patterns. For example, "ptinfo ◊ lab ◊ lab^.KEY." is a 3-sequence. We call these patterns, "general" patterns because they only provide a general idea about usage.From the set of 3,462 general patterns, we extracted those that occurred in 1% or more of the user sessions. This gave us a set of 418 patterns. Table 2 contains the breakdown of n-sequences in this set.

The 418 general patterns were further classified as simple (n=96) and complex (n=322). Simple patterns are those that contain only data types that cannot be further expanded such as "ptinfo", "lab", and "rad". Table 3 contains some simple patterns.

We expanded 146 of the complex patterns to obtain "detailed" patterns, which are at a finer level of granularity. Complex patterns consist of both data types that can be further expanded such as "lab^.KEY." and those that cannot such as "card" and "path". To obtain the detailed patterns, we expand ".KEY." in each pattern to the actual result name. For example, "lab^.KEY." could be expanded to "lab^Basic Metabolic Panel". Table 4 contains some complex patterns and the number of detailed patterns associated with them. Table 5 contains partial expansions for three of the complex patterns.

Table 1: Data Type Statistics

Data Type	Definition	Frequency
lab^.KEY.	patient-specific laboratory result	24.6%
lab	laboratory	8.4%
rad^.KEY.	patient-specific radiology result	5.2%
rad	radiology	3.8%
path^.KEY.	patient-specific pathology result	1.9%
path	pathology	1.6%
card^.KEY.	patient-specific cardiology result	1.6%
card	cardiology	1.2%

Table 2: Number of n-sequences with frequency >=1.0%

n	# of sequences	n	# of sequences
2	100	7	32
3	89	8	20
4	63	9	17
5	45	10	15
6	37	TOTAL	418

Discussion

The analysis we performed identified patterns of usage in a patient-record system. These sequential patterns represent one aspect of the patient-specific information needs of clinicians.

While many of the patterns found are intuitive, analysis of the results reveals those that are not as intuitive. For example, we expect that viewing of liver function test results would follow viewing of an abdominal ultrasound report as found in Table 5.

ptinfo → lab → lab ∧^Coag Profile → lab ∧ ABC → lab ∧ Basic Metabolic Panel → lab ∧^Coag Profile
ptinfo → lab → lab ∧ .Key.→ lab ∧ .Key.→ lab ∧ .Key.→ lab ∧ .Key.

ptinfo → card → card ∧^Transthoracic Echocardiography → card ∧^Transthoracic Echocardiography →
lab ∧ 72-hour lab update → lab ∧ 7-day lab update → lab → lab ∧ Basic Metabolic Panel →
rad → rad ∧ Computerized Axial Tomography of Head without Contrast → rad ∧ X-Ray of Chest, 2 Views → rad ∧ X-Ray of Swallowing Function with Cine
ptinfo → card → card ∧^.KEY.→ card ∧^.KEY. → lab ∧ 72-hour lab update → lab ∧ 7-day lab update → lab → lab ∧ .KEY. → rad → rad ∧^.KEY.→ rad ∧^.KEY. → rad ∧^.KEY.

Figure 2 - Two user sessions at fine level and intermediate level (in italics) of granularity. The data type «ptinfo» indicates that a patient record is being accessed. «.Key.» is used to indicate that a result was accessed, but does not specify the type

Table 3: Simple Patterns

Simple Patterns	Frequency
ptinfo → lab	68.6%
ptinfo → rad	19.1%
ptinfo → lab^12-hour lab update	10.9%
ptinfo → path	8.7%
ptinfo → signout	5.2%
ptinfo → card	4.1%
lab^spreadsheet^WEBCIS BLOOD COUNT DISPLAY → lab^spreadsheet^WEBCIS BASIC METABOLIC DISPLAY	4.1%
ptinfo → lab^72-hour lab update	3.5%
lab^spreadsheet^ WEBCIS BASIC METABOLIC DISPLAY → lab^spreadsheet^WEBCIS HEPATOBILIARY DISPLAY	2.2%
rad → lab	2.2%

We also could predict that users would select thyroid function tests after viewing a thyroid ultrasound report. However, the choice of "Lipid Profile" may not be as predictable, until one considers that thyroid abnormalities often lead to derangements in lipid metabolism.

We learned which functions are more popular and identified patterns, which convey how users access patient data in the system. These findings can be stored in a knowledge base for use by clinical information systems such as WebCIS and PalmCIS.

WebCIS

Several types of enhancements can be made to WebCIS using the results of the CIS log analysis. First, the data type statistics indicate which WebCIS functions are more popular. This knowledge can be used to make these functions more accessible to users, thereby saving them time when traversing the system. For example, we learned users most frequently click on “Laboratory”, “Radiology”, “Pathology”, and “Cardiology” to view the respective listings of results. One enhancement we can make is to modify the leftmost frame containing the menu to list the functions by frequency of use.

Currently, users must search WebCIS to find the patient data of interest. This may involve an unnecessary amount of page clicks and time. The knowledge base can be used to anticipate what the user may want to view next and provide links to those data.

These context-sensitive “shortcuts” can be generated for each page to provide quicker access to needed information [12].

In Table 5, we have a rule that shows that when a radiology result for “Diagnostic Ultrasound of Abdomen” is viewed, the laboratory result viewed next may be “Hepatic Function Panel”, “Basic Metabolic Panel”, or ”ABC”. When users are viewing these types of radiology results, we can now add a shortcut list containing these types of laboratory results. The user can select the appropriate shortcut and quickly obtain the results desired.

Alternatively, the knowledge base can be used to generate a report containing a summary of results. Rather than access results individually, we can anticipate the sequence of results users desire and display them together. WebCIS currently contains lab summaries and lab updates, which aggregate various laboratory results. Users can also create patient record summaries by specifying the time period and types of results to view.

From the simple patterns obtained, we learned that when users first access a patient’s record, they view laboratory results (68.6%), radiology results (19.1%) or pathology results (8.7%).

Currently, users would have to click on each of these departments to see the listings (or create a summary themselves). Using the knowledge base, we could automatically generate a report containing these three listings for display.

PalmCIS

PalmCIS (Palm-based Clinical Information System) provides access to patient data via a wireless, handheld device [13]. Because of the limitations of this environment, the goal is to present only patient data that are needed in an efficient manner.

PalmCIS currently contains a subset of data that can be found in WebCIS. The initial design was based on feedback from a set of clinicians. Upon selection of a patient, users are presented with the patient’s report, which contains a summary of laboratory, radiology and cardiology results from the past two days. Users then have the option to view result details and results from previous days as well as well as pharmacy reports, discharge summaries and sign-out notes via a menu following the report.

Table 4: Complex Patterns

Complex Patterns	Frequency	Number of Detailed Patterns
ptinfo → lab → lab^.KEY. → lab^.KEY. → lab^.KEY. → lab^.KEY. → lab^.KEY.	15.2%	274,568
lab^.KEY. → rad → rad^.KEY. → rad^.KEY.	1.9%	29,875
rad^.KEY. → card → card^.KEY.	1.8%	1,411
lab^.KEY. → path → path^.KEY.	1.8%	4,659
rad^.KEY. → lab → lab^.KEY. → lab^.KEY. → lab^.KEY. → lab^.KEY. → lab^.KEY.	1.5%	42,435
rad^.KEY. → path → path^.KEY.	1.5%	4,938
path^.KEY. → lab → lab^.KEY.	1.2%	10,066
ptinfo → rad → rad^.KEY. → lab → lab^.KEY.	1.2%	4,011
path^.KEY. → rad → rad^.KEY.	1.2%	15,360
ptinfo → rad → rad^.KEY. → rad^.KEY. → rad^.KEY. → rad^.KEY.	1.0%	24,443

Table 5: Detailed Patterns

Detailed Patterns			Frequency
ptinfo → rad → rad^.KEY. → lab → lab^.KEY.			
		Hepatic Function Panel	19.3%
ptinfo → rad → Diagnostic Ultrasound of Abdomen → lab		Basic Metabolic Panel	7.9%
		ABC	4.1%
		Thyroid Panel B	43.1%
ptinfo → rad → Diagnostic Ultrasound of Thyroid → lab		T3	11.3%
		Lipid Profile	5.0%
path^.KEY. → rad → rad^.KEY.			
		MRI of Brain with and without Contrast	31.8%
Brain biopsy → rad		CT Guidance of Stereotactic Localization	10.2%
		CT of Head without Contrast	9.1%
		Diagnostic Ultrasound of Abdomen	31.8%
Gallbladder, cholecystectomy → rad		X-Ray of Chest, 2 Views	8.9%
		Cholangiogram in Surgery	5.0%
rad^.KEY. → card → card^.KEY.			
		12-Lead Electrocardiogram	45.0%
CT of Head without Contrast → card		Transthoracic Echocardiography	39.9%
		Transesophageal Echocardiography	4.4%
		Transthoracic Echocardiography	38.9%
Diagnostic Ultrasound of Abdomen → card		12-Lead Electrocardiogram	34.5%
		Cardiac Catheterization	9.5%

Due to the simpler design and functionality of PalmCIS, the statistics and knowledge base have a different impact. The data type statistics can be used to determine which functions to incorporate into PalmCIS, while the knowledge base can identify what to include in the initial patient report and menu. Additionally, shortcut lists can be included that provide direct access to needed results. Figure 3 contains screenshots of shortcuts in PalmCIS.

Future Work

We focused on finding sequential patterns of data types, which just convey what types of results were viewed sequentially in WebCIS. Next steps include looking at the time associated with these results and identifying patterns based not only on result type but also timestamp.

The statistics and patterns we have presented apply to all WebCIS users. From classification, we can determine if user characteristics (e.g. role, specialty) and possibly patient characteristics (e.g. age, gender, diagnosis) have an influence on statistics and patterns. We may also discover that the time of day the user is accessing WebCIS has an impact. This information as well as results from association rule generation can be added to the knowledge base to further customize menus, shortcut lists, and patient reports.

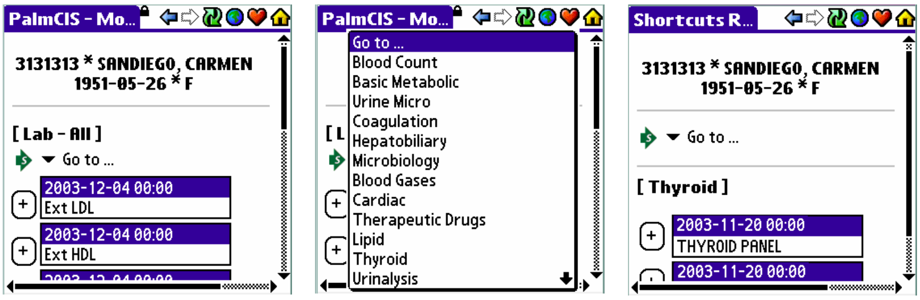

Figure 3 - Shortcuts in PalmCIS. The user chose to view a list of all laboratory results and then selected a shortcut to Thyroid results

We have looked at using CIS log analysis to learn about patient-specific information needs from usage of WebCIS, a Web-based clinical information system that provides access to patient information. We are now interested in determining how accurate this method is for identifying user needs and the impact of this knowledge. Next steps include evaluating the patterns obtained from the analysis to determine their accuracy. We are currently adding shortcut lists to PalmCIS as discussed. Once this is completed, we can evaluate the impact and usefulness of these shortcuts.

Conclusion

CIS log analysis specializes in finding patterns in sequential clinical data. We have used this method to automatically identify the patient-specific information needs of thousands of WebCIS users from system log files. The WebCIS log files contain a rich amount of information about users. Each field in a log record provides a different view of the users. We obtained basic statistics on each of the fields and used the data type field to identify usage patterns. This knowledge can be used to reorganize menus, to dynamically create shortcuts to needed information, and to generate patient reports in real time for clinical information systems dealing with patient-specific data.

Acknowledgments

This work has been supported in part by National Library of Medicine Training Grant NO1-LM07079. The authors thank Soumitra Sengupta and Richard Gallagher for their assistance in obtaining the WebCIS log files.

References

[1] Smith R. What clinical information do doctors need? *BMJ* 1996 Oct 26;313(7064):1062-8.

[2] Hripcsak G, Cimino JJ, Sengupta S. WebCIS: large scale deployment of a Web-based clinical information system. *Proc AMIA Symp* 1999;:804-8.

[3] Cimino JJ, Socratous SA, Clayton PD. Internet as clinical information system: application development using the World Wide Web. *J Am Med Inform Assoc* 1995 Sep-Oct;2(5):273-84.

[4] Gallagher RJ, Sengupta S, Hripcsak G, Barrows RC, Clayton PD. An audit server for monitoring usage of clinical information systems. *Proc AMIA Symp* 1998;:1002.

[5] Sengupta S, Gallagher R, Hripcsak G, Wu M, Barrows RC, Clayton PD. Searching for patterns in an audit log of clinical information access. *Proc AMIA Spring Congress* 1997;:55.

[6] Haigh S and Megarity J. Measuring Web site usage: log file analysis. Network Notes #57. 1998. Available at: http://www.nlc-bnc.ca/publications/1/p1-256-e.html

[7] D'Alessandro MP, D'Alessandro DM, Galvin JR, Erkonen WE. Evaluting overall usage of a digital health sciences library. *Bull Med Libr Assoc* 1998 Oct;86(4):602-9.

[8] Gardois, P. Measuring the performance of a biomedical digital library: web site, e-journals and databases. *EAHIL Workshop* 2001;:

[9] Rozic-Hristovski A and Hristovoski D. Users' information-seeking behavior on a medical library Website. *J Med Libr Assoc* 2002 Apr;90(2):210-217.

[10]Westbrook JI, Gosling AS, Coiera EW. Do clinicians use online evidence to support patient care? A study of 55,000 clinicians. *J Am Med Inform Assoc* 2004 (in press)

[11] Chen ES, Cimino JJ. Automated discovery of patient-specific clinician information needs using clinical information system log files. *Proc AMIA Symp* 2003;:145-149.

[12] Chen ES, Hripcsak G, Patel VL, Sengupta S, Gallagher RJ, Cimino JJ. Automated identification of shortcuts to patient data for a wireless handheld clinical information system. *Proc AMIA Symp* 2003;:809.

[13] Chen ES, Mendonca EA, McKnight LK, Stetson PD, Lei J, Cimino JJ. PalmCIS: a wireless handheld application for satisfying clinician information needs. *J Am Med Inform Assoc*. 2004 Jan-Feb;11(1):19-28.

Address for correspondence

Elizabeth S. Chen
Department of Biomedical Informatics
Columbia University
622 W 168th Street
Vanderbilt Clinic, 5th Floor
New York, NY 10032
Email: liz.chen@dbmi.columbia.edu