

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



Radiology Data Mining Applications using Imaging Informatics

Richard Chen MD, MBA¹, Pattanasak Mongkolwat PhD²
and David Channin MD²

¹Ohio State University College of Medicine, Columbus, OH

²Northwestern University Feinberg School of Medicine, Chicago, IL
USA

1. Introduction

The radiology department is a unique department within the healthcare enterprise. It has its roots within technology and is a naturally information-rich area within which data can be mined, analyzed and used to improve departmental operations. The recent migration of many healthcare enterprises to PACS (picture archiving and communication systems) helps to facilitate this trend. This chapter will provide an overview of the various technologies in the radiology department that allow for data-mining. Case-examples utilizing these technologies are then discussed in detail.

2. PACS and DICOM

PACS are computer distribution networks that are dedicated to the storage and retrieval of medical images. The systems started as simple image management systems which have currently evolved to include images, voice, text, medical records, and video recordings (Huang et al, 2005). Numerous studies have reported productivity gains associated with PACS implementations (Mackinnon et al., 2008). Fundamental to PACS is the display of images using Digital Imaging and Communications in Medicine (DICOM) headers.

DICOM is a universally used format for medical image storage. The standard was originally created by the National Electrical Manufacturers Association (NEMA) and the American College of Radiology (ACR) for medical image archiving and communication (NEMA, 2008, Horii et al., 1992). It is currently maintained by the DICOM Standards Committee. It is important as it guarantees a minimum level of compatibility between hardware and software publishers from a number of different manufacturers. DICOM encapsulates dataset objects which accompany the image. These dataset objects contain much information about the associated image(s). DICOM files contain a header which specifies information related to patient name, medical record number, image type (x-ray, CT scan, MRI, ultrasound, etc), and image dimension. The files then also contain all of the image data. This differs from other formats which store the header and image data in separate files. DICOM also allows storage space for more elaborate data constructions, from which an innumerable amount of information may be gleaned. This includes information such as

Source: Data Mining in Medical and Biological Research, Book edited by: Eugenia G. Giannopoulou,
ISBN 978-953-7619-30-5, pp. 320, December 2008, I-Tech, Vienna, Austria

image transit information across the PACS system marked with timestamps, or radiologist view and dictation times (NEMA DICOM, 2008).

3. Health level 7 and integrating the healthcare enterprise

In addition to DICOM standardization for PACS platforms, there are multiple efforts underway to standardize and improve the way with which information is exchanged within the healthcare industry. The goal is to develop semantic language structures to meet the demands and challenges of inter-operability between existing and future generations of health information systems (Lopez et al., 2008, Blobel et al., 2006). One such initiative is Health Level 7 (HL7), a volunteer, not-for-profit organization which provides a framework to share electronic health information (HL7, 2008). The scope of HL7 includes standards for the exchange of clinical data in all settings – it allows the exchange of information through a generalized reference informational model, via various data types, and through the use of decision support trees. It includes provisions for security, XML data exchange, and general electronic health record use (Hammond, 2003). Derivations of the reference informational model also allow inclusion of clinical documents, such as the ANSI-approved Clinical Document Architecture (CDA), a document markup standard that specifies the structure and semantics of clinical documentation (Dolin et al., 2001).

Another initiative to facilitate healthcare data exchange is that of the IHE (Integrating the Healthcare Enterprise), a collaboration between healthcare professionals and industry members to improve data interoperability, focusing on the development of global integration profiles and a clear implementation path using existing medical standards such as DICOM and HL7, while encouraging members to uphold compatibility (IHE, 2008). IHE is important because it approaches tasks at a much higher level, taking care not to specify specific roles for specific applications, but rather defining a set of generic “actors” and a set of roles that these actors must play to successfully accomplish a given task (Channin, 2000). Systems supporting IHE have been shown to enable healthcare providers to use information more effectively in providing better and more-informed care when managing patients ((Lian et al., 2006). Its success lies in its adoption by industry professional organizations representing both buyers and vendors (Hussein et al., 2004).

4. Case example: RadMonitor

4.1 Introduction.

This section describes a case example, RadMonitor. The tool was designed within our department as a platform-independent web application designed to help manage the complexity of information flow within a healthcare enterprise. The system eavesdrops on HL7 traffic and parses statistical operational information into a database. The information is then presented to the user as a treemap – a graphical visualization scheme that simplifies the display of hierarchical information. While RadMonitor has been implemented for the purpose of analyzing radiology operations, its XML backend allows it to be reused for virtually any other hierarchical dataset.

4.2 Technologies.

The RadMonitor design involves a traditional three-tier architecture consisting of the database, server and client (Figure 1). Radiology operations information flows from an HL7 data feed to a MySQL relational database. Clients then interact with this data through a JavaScript enabled web browser and a server-side processing script.

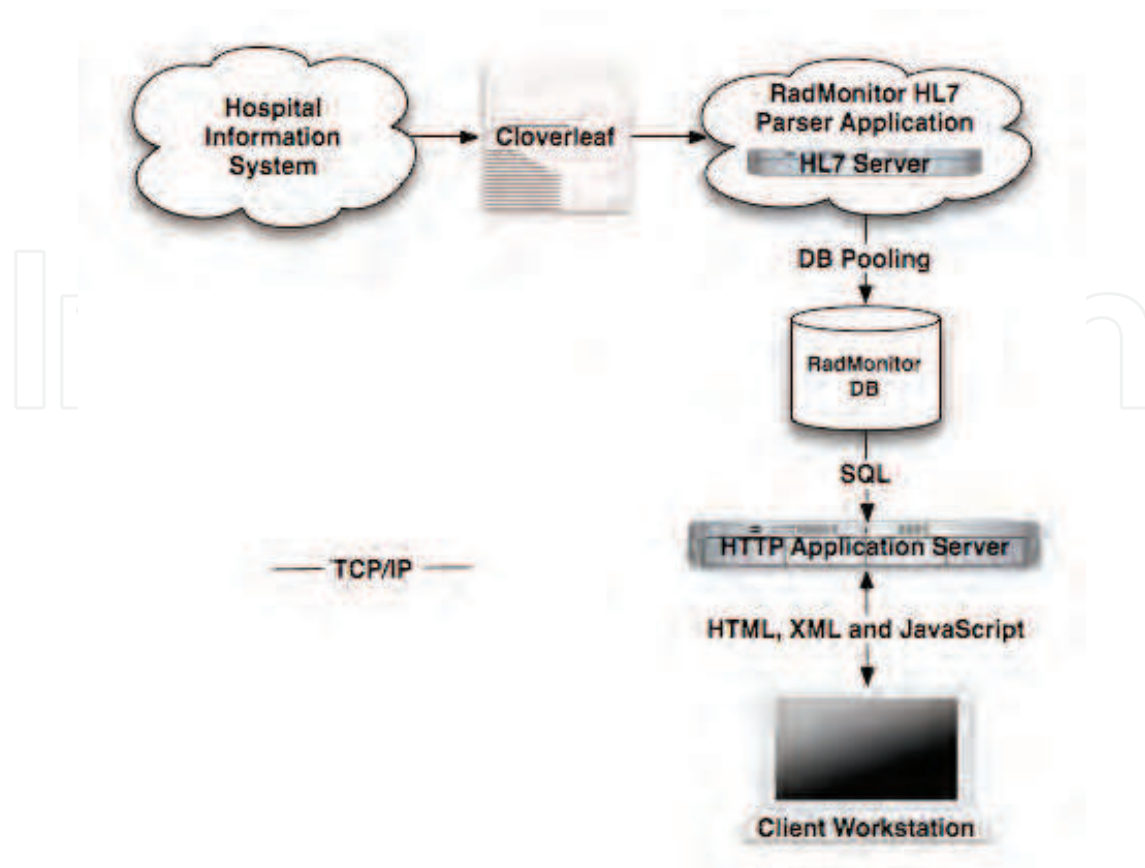


Fig. 1. RadMonitor implementation design.

In the backend, an HL7 application on the server receives order message (ORM) and result message (ORU) data feeds from Quovadx, formerly known as Cloverleaf. It is used to parse the HL7 messages, and allows us to map, route and extract data from the hospital information system. Statistical information relevant to radiology operations management is stored in the RadMonitor database. Database entries include fields that represent order status changes, and start and stop dictation and transcription time events.

The information in the database is available to the client to query and retrieve. This is done through a server-side processing script written in PHP. All of the exchanges between the server and client are done using HTML, XML and JavaScript. While these technologies have been around for some time, RadMonitor makes extensive use of a relatively new twist of the technology - the XmlHttpRequest object.

4.3 Implementation.

RadMonitor was implemented to improve the delivery and utility of the treemap construct through a web platform. Our goal was to build a lightweight, extensible and standards-based solution, complete with the interactivity and responsiveness expected of a desktop application. Application interactivity was a key component in our design, and the use of AJAX allowed us to dramatically improve the user experience by deviating from the traditional web interaction model (Figure 2).

Initially, only HTML is rendered in the client web browser. This acts as an empty layout template, upon which subsequent AJAX queries fill respective sections of the page. The first section to be loaded is the treemap itself.

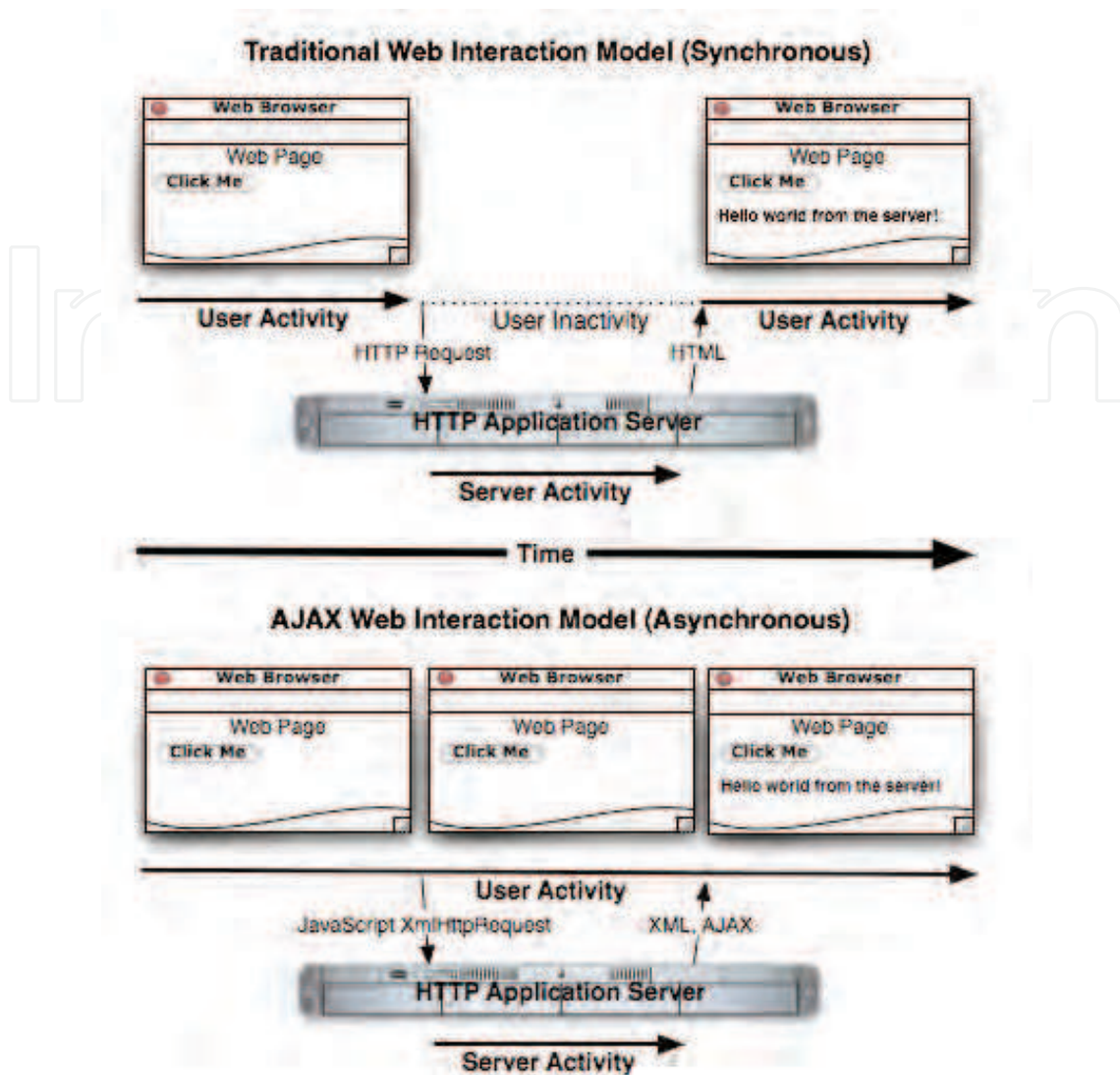


Fig. 2. The asynchronous AJAX user model provides for a user experience that is continuous without user inactivity as a result of time lags associated with page refreshes in a synchronous model.

The treemap is the centerpiece of the application (Figure 3). The three treemaps that RadMonitor currently supports are Orders, Radiologist and Staff. These selections are made using the radioboxes in the upper right corner of the page. The Radiologist treemap is a representation of the radiologists' average dictation time. Information is divided into a hierarchy of modalities, and radiologists within a modality. The size of an individual radiologist's rectangle is related to the number of studies that the radiologist has dictated. Correspondingly, the size of a modality's rectangle is indicative of the total number of studies dictated within that modality. The color and color gradient of a radiologist's rectangle is a measure of the average time that the radiologist has spent dictating exams as compared to the modality average. Green represents an average dictation time faster than the modality average, and red indicates times slower than the average. A similar hierarchical breakdown is applied to the Staff treemap, which displays information based on the staffs' average transcription time.

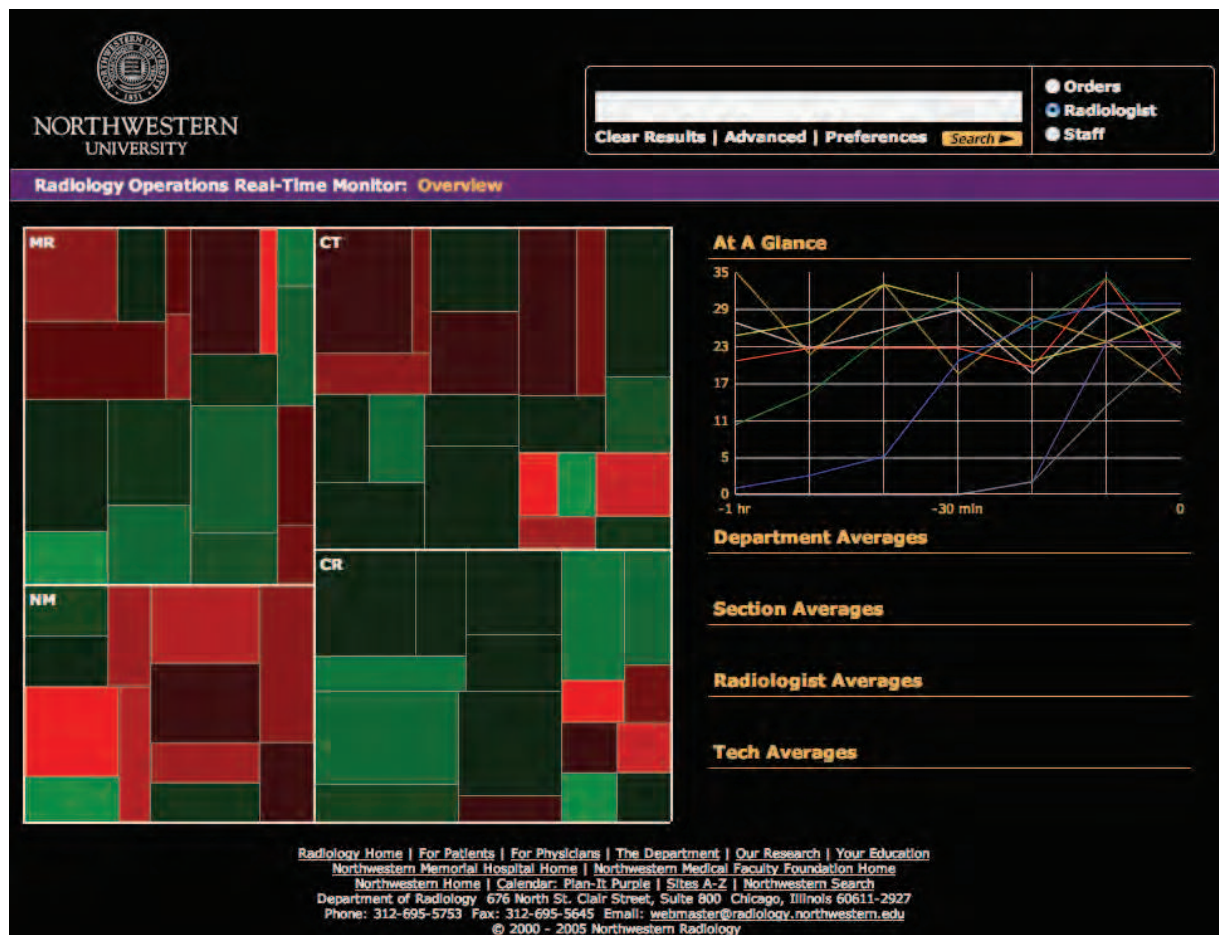


Fig. 3. RadMonitor application showing treemap with radiologists' comparative average dictation time represented by color and films read represented by size.

4.4 Discussion.

The use of these parameters results in a treemap that conveys information about the real-time operational statistics of the radiology department. This information is intuitively represented, and allows us to readily answer questions such as: Which modality reads the most number of exams? The least? How are the dictation times different between modalities? How about within a modality?

Information is the lifeblood of any business and healthcare is no exception. No single information system can manage the entirety of a healthcare enterprise of any significant size. RadMonitor is a tool designed to help manage the complexity of information flow within the hospital network. As with the dashboard concept, its utility lies in its use of proprietary or open data standards to interact and interface with other hospital information systems.

HL7 and DICOM standards are critical to this process. Specifically, the data feeds are instrumental for the communication of non-image and image based information, respectively. Furthermore, both HL7 and DICOM data feeds can serve as a source of valuable information about the status of operations in a department. A surprising amount of analytical information is contained within these messages. In addition to being vital to

clinical and research activities, this information can be monitored, in real-time, to provide immediate decision support for administrative and management activities.

5. Case example: PACSMonitor

5.1 Introduction.

PACSMonitor is a tool designed to help manage the complexity of radiology information flow within the hospital network. As with the dashboard concept, its utility lies in its use of proprietary or open data standards to interact and interface with other hospital information systems. As with the above example, PACSMonitor mines HL7 and DICOM data for valuable information about the status of operations in the department. This information can be monitored, in real-time, to provide immediate decision support for administrative and management activities.

5.2 Technologies.

The PACSMonitor implementation involves a traditional three-tier architecture consisting of the database, server and client. The software was built to seamlessly interface with typical commercial PACS distributions. The stability of a PACS system with minimal unexpected PACS downtime is a fundamental concern for all hospitals. In order to maximize server load and minimize downtime and instability without adversely impacting system performance, PACSMonitor mines PACS information related to patient studies, examination acquisition studies and so forth during a daily off-peak cycle and stores this information in a separate pre-parsed database exclusively dedicated to PACSMonitor. In the backend, an HL7 application on the server receives order message (ORM) and result message (ORU, RAD-28) data feeds from Quovadx, formerly known as Cloverleaf. The ORM messages are based on IHE technical framework transactions RAD-1, RAD-2, RAD-3, RAD-4, RAD-6, RAD-7, RAD-10, RAD-13, RAD-20, RAD-21, RAD-42, and RAD49. These HL7 messages are parsed according to our application needs, allowing us to transparently map, route and extract data from the hospital information system. Statistical information relevant to radiology operations management is stored in the PACSMonitor database.

The information in the database is immediately available for clients to query. All client interaction with the web application is done via server-side processing scripts written in PHP. The exchanges between the server and client are all based on web standards using HTML, XML and JavaScript.

5.3 Implementation.

The main screen of PACS monitor is divided into a right frame comprising of various user reports and a left main window which houses both the report builder and the visualizations (Figure 4). Creating a report starts by choosing the database table from which the user wishes to query information. These table names are dynamically populated by inspection of the database and including all tables that are identified as containing statistical data. Next, the user has the option of choosing the type of report he wishes to view, including "Table", "Graph", or "Treemap". Each of these report types include options specific to their display. A table report, for instance, allows users to select which columns that they would like to include. The graph report allows users to specify the x and y axis, specific colors, as well as

the type of graph (Figure 5). The treemap visualization allows users to select both the color and size of the data representations. Users are further able to customize the reports that they create using the “Match the following rules” option. This option allows users to fine-tune the data that they query from the database. Creating the report causes the new report to show up in the right frame.

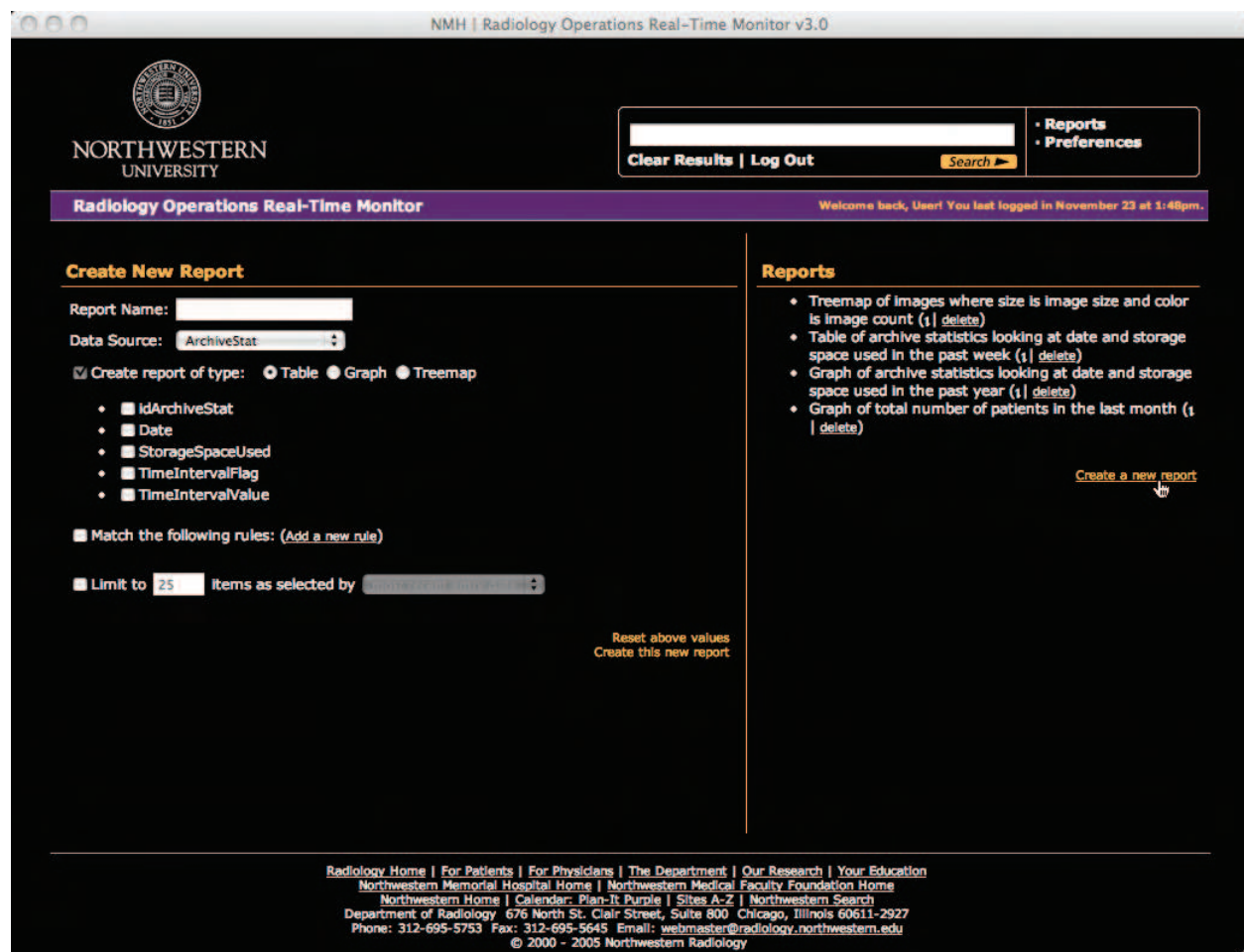


Fig. 4. This is a screenshot of the default view of the application. In the right frame, you can select an existing report to view or create a new report. The left frame shows an example of creating a new report.

The report builder interface simplifies the workflow by which users choose the information that they want to obtain. It offers standardized templates from which to choose and construct reports. Each report is stored on the database as a dynamically constructed SQL query. At the same time, the interface offers flexibility and customization for power users via “smart” selections that allow users to apply selective filters to the data queried from the database (Figure 6). These filters modify the SQL query to include modifiers such as “where”, “limit”, or “sort by”. It is an example of the power of user-specified and user-constrained data, similar to that of the “smart playlist” concept in Apple’s iTunes software. Users are thus virtually unlimited in the number of ways by which they can query specific data from the backend servers. Clicking on a report to view triggers the SQL query for the requested information, and the resulting dataset is sent to the report viewer script to render.

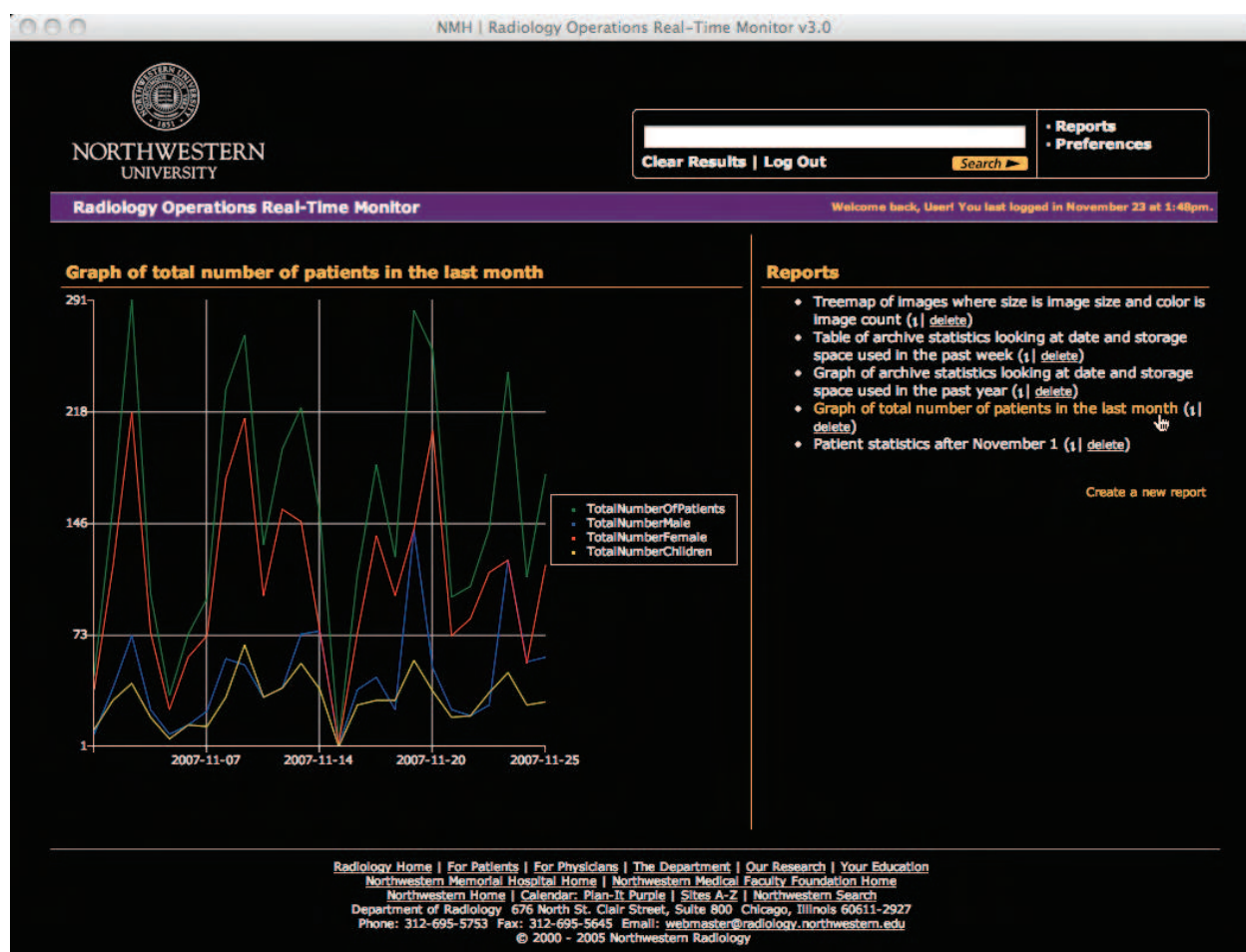


Fig. 5. Many different data combinations are possible. This graph shows the total number of patients within the past month.

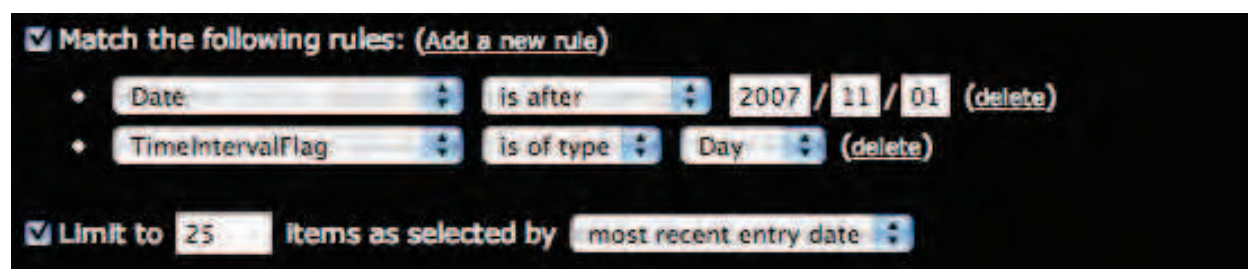


Fig. 6. Users are able to provide precise modifiers on the type of data that they want to see.

5.4 Discussion.

The gigabytes of radiologic imaging information that is transferred daily within a hospital enterprise PACS installation is associated with significant metadata, or data describing data. These bits, flags, dates, and numbers can be obtained by eavesdropping on HL7 traffic. By parsing and running statistical calculations on specific items of interest, this metadata serves

as a powerful metric for analyzing and interpreting real-time radiology operational information.

PACSMonitor provides web-based operational reports of underlying PACS metadata. Instead of mining this information directly from the PACS, an activity that could threaten its stability and performance, we make periodic transfers to a separate database and then provide tools to the users to visualize this information. Examples of daily reports include “studies not yet completed within 24 hours” or “cancelled studies containing valid images”. Longer-term monthly reports may describe the “rejected image rate” or “examinations without a procedure description”. The ease with which users can access these metrics makes this a powerful tool to facilitate interpretation of daily and long-term operations in a data-driven fashion.

6. References

- Blobel BG, Engel K, Pharow P. (2006). Semantic interoperability--HL7 Version 3 compared to advanced architecture standards. *Methods Inf Med.* 2006;45(4):343-53.
- Channin DS. (2000). M:I-2 and IHE: integrating the healthcare enterprise, year 2. *Radiographics.* 2000 Sep-Oct;20(5):1261-2.
- Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo Shvo A. (2006). HL7 Clinical Document Architecture, Release 2. *J Am Med Inform Assoc.* 2006 Jan-Feb;13(1):30-9. Epub 2005 Oct 12.
- Hammond WE. (2003). HL7--more than a communications standard. *Stud Health Technol Inform.* 2003;96:266-71.
- Health Level 7 (HL7). (Last accessed June 2008). <http://www.hl7.org>
- Horii SC, Bidgood WD Jr. (1992). PACS Mini Refresher Course: Network and ACR-NEMA Protocols. *Radiographics.* 1992 May;12(3):537-48.
- Huang HK. (2005). *PACS and Imaging Informatics: Basic Principles and Applications.* Hoboken, NJ: Wiley, 2004. ISBN 0-471-25123-2.
- Hussein R, Engelmann U, Schroeter A, Meinzer HP. (2004). Implementing a full-feature PACS solution in accordance with the IHE technical framework: the CHILI approach. *Acad Radiol.* 2004 Apr;11(4):439-47.
- Integrating the Healthcare Enterprise (IHE). (Last accessed June 2008). <http://www.ihe.net>
- Lian JD, Lin IC, Wu HC. (2006). Case report: Taiwan's experience in adopting IHE technical framework to integrate systems. *Stud Health Technol Inform.* 2006;122:877.
- Lopez DM, Blobel BG. (2008). A development framework for semantically interoperable health information systems. *Int J Med Inform.* 2008 Jul 11. [Epub ahead of print].
- Mackinnon AD, Billington RA, Adam EJ, Dundas DD, Patel U. (2008). Picture archiving and communication systems lead to sustained improvements in reporting times and productivity: results of a 5-year audit. *Clin Radiol.* 2008 Jul;63(7):796-804. Epub 2008 Mar 25.

National Electrical Manufacturers Association (NEMA). (Last accessed June 2008).
<http://medical.nema.org/>

National Electrical Manufacturers Association DICOM (NEMA DICOM). (Last accessed
June 2008). <ftp://medical.nema.org/medical/dicom/2008/>

IntechOpen

IntechOpen



Data Mining in Medical and Biological Research

Edited by Eugenia G. Giannopoulou

ISBN 978-953-7619-30-5

Hard cover, 320 pages

Publisher InTech

Published online 01, November, 2008

Published in print edition November, 2008

This book intends to bring together the most recent advances and applications of data mining research in the promising areas of medicine and biology from around the world. It consists of seventeen chapters, twelve related to medical research and five focused on the biological domain, which describe interesting applications, motivating progress and worthwhile results. We hope that the readers will benefit from this book and consider it as an excellent way to keep pace with the vast and diverse advances of new research efforts.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Richard Chen, Pattanasak Mongkolwat, and David Channin (2008). Radiology Data Mining Applications Using Imaging Informatics, Data Mining in Medical and Biological Research, Eugenia G. Giannopoulou (Ed.), ISBN: 978-953-7619-30-5, InTech, Available from:
http://www.intechopen.com/books/data_mining_in_medical_and_biological_research/radiology_data_mining_applications_using_imaging_informatics

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2008 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen