Application of Information Technology ■

Problem-oriented Prefetching for an Integrated Clinical Imaging Workstation

ALEX A.T. BUI, PHD, MICHAEL F. MCNITT-GRAY, PHD, JONATHAN G. GOLDIN, MD, PHD, ALFONSO F. CARDENAS, PHD, DENISE R. ABERLE, MD

Abstract Prefetching methods have traditionally been used to restore archived images from picture archiving and communication systems to diagnostic imaging workstations prior to anticipated need, facilitating timely comparison of historical studies and patient management. The authors describe a problem-oriented prefetching scheme, detailing 1) a mechanism supporting selection of patients for prefetching via characterizations of clinical problems, using multiple data sources (picture archiving and communication systems, hospital information systems, and radiology information systems), classifying patients into cohorts on the basis of their medical conditions (e.g., lung cancer); and 2) prefetching of multimedia data (imaging, laboratory, and medical reports) from clinical databases to enable the viewing of an integrated patient record. Preliminary evaluation of the prefetching algorithm using classic information retrieval measures showed that the system had high recall (100 percent), correctly identifying and retrieving data for all patients belonging to a target cohort, but low precision (50 percent). A key finding during testing was that the recall of the system was increased through the use of multiple data sources (compared with one data source), because of better patient descriptors. Medical problems and patient cohorts were more specifically defined by combining information from heterogeneous databases.

■ **J Am Med Inform Assoc.** 2001;8:242–253.

To accurately evaluate the temporal evolution of a medical condition, access to historical data is required to facilitate comparison of a patient's current status with findings from past examinations. All available information in a patient's medical record—such as laboratory reports, admission/discharge/transfer (ADT) reports, imaging findings—should be presented in an integrated fashion to provide a comprehensive understanding of the patient's medical history. Ideally, all the patient data (both past and present) is made quickly available to the clinicians involved in the patient's care.

Affiliation of authors: University of California at Los Angeles (UCLA), Los Angeles, California.

This work was funded by grant PO1-CA51168-07 from the National Institutes of Health (NIH).

Correspondence and reprints: Alex Bui, PhD, Telemedicine Division, UCLA Department of Radiology, 924 Westwood Boulevard, Suite 420, Los Angeles, CA, 90024; e-mail: <buia@cs.ucla.edu>.

Received for publication: 10/15/98; accepted for publication: 1/18/01.

Today's picture archiving and communication systems (PACSs) make it possible to readily retrieve images from prior imaging examinations. Given the appropriate bandwidth (e.g., 100 base-T Ethernet, asynchronous transfer mode), image sets can be quickly transferred from a PACS to a diagnostic workstation. For example, multi-sequence helical CT (computed tomography) or MR (magnetic resonance) scans, which can range in size from 2 to 20 MB, can normally be sent over such networks in less than 10 seconds; even a few seconds, however, are unacceptable to most clinicians.[1,2] In addition, such retrieval performance is under ideal conditions, and other factors must be considered, such as the restoration of older images from tertiary storage.

Most PACS are based on a hierarchical storage structure, with past images archived on slower media, such as optical disc and tape. As such, images from older examinations require additional time to locate and restore, thus incurring a further delay in retrieval time; for instance, retrieval from tape archives averages 15 minutes.[3] For patients with extensive imaging histories, multiple archived files must be fetched, contributing further to the overall retrieval time. To address this issue, prefetching algorithms are used to fetch historical data to workstations before the anticipated time of need, removing delays due to the process of restoring archived data.

At UCLA, the development of a multimedia imaging workstation for oncology patients has been the impetus for designing a new prefetching system for managing information retrieval for specific groups of patients. In addition to utilizing "standard" prefetching information, the prefetching algorithm takes advantage of further sources of patient information (e.g., hospital and laboratory reports) to 1) classify individual patients into specific cohorts, directing relevant data to appropriate workstations and specialists (e.g., classifying a patient into a lung cancer cohort and sending the data to a thoracic oncologist's workstation); 2) initiate retrieval of data from PACSs and other patient databases (the hospital information system and radiology information system); and 3) trigger external processing of retrieved data (e.g., image processing routines) so that additional information is available for clinicians at the time of review.

By incorporating all electronic information—for which image data are but one component—this "problem-oriented" prefetching strategy makes possible an integrated review of patient data. In addition, as databases can be queried to retrieve patient cohorts with similar medical conditions, outcome research addressing selective questions can be supported (e.g., "Identify all nonsmoking females between the ages of 35 and 50 years with multifocal bronchioloalveolar cell carcinoma").

This paper describes the design and implementation of our prefetching system and is organized as follows. First, we give the background and motivation for our project. Second, the prefetching algorithm is explained and the implementation of its components— a rule language, a user interface for rule editing, and a clustering mechanism for grouping similar patients together—are detailed. An initial evaluation of the system using standard information retrieval measures is also described.

Background and Motivation

The escalating trend to integrate all aspects of the digital medical environment has led to interest in developing multimedia computerized patient records.[4,5] For example, the NUCLEUS project[6] created integrated multimedia records using a hypermedia metaphor. Similarly, the hypercard environment was used in developing a multimedia clinical record for radiotherapy,[7] and an early version of the MARS Image Engine was used to construct a multimedia medical database.[8,9]

In a joint collaboration among medical oncologists, radiologists, and computer scientists, a multimedia database application, the Oncology Imaging TimeLine (OITL)[10,11] was developed, serving as an integrated medical record for oncology patients. Our goal was to develop a shared interface that would expressly aid the information exchange between radiologists and clinicians caring for patients with complex medical diseases. At our institution, patient rounds are performed at specific times according to specialty (e.g., oncology, urology). Oncology rounds, for instance, have participation from medical and surgical oncologists, radiologists, and house staff. OITL establishes a multimedia database, supporting the viewing of an integrated patient record in terms of chronologies, making available oncologic patient data in a convenient manner for review by this team.

Unlike past efforts focusing on temporal visualization of patient records,[12–14] OITL also addresses the underlying database foundation for the user interface, explicitly managing time-based data using stream data constructs.[15,16] OITL's threefold role is summarized as follows:

- As in other medical centers, a large portion of our patients' medical histories are stored electronically in multiple databases: the hospital information system (ADTs, clinic notes, laboratory reports, pathology reports), the radiology information system (radiology reports), and PACS (digital images). OITL automatically collects and integrates information about cancer patients from these distributed data sources.
- OITL executes data extraction routines (e.g., image processing), providing additional information to clinicians during the review process.
- OITL re-arranges the totality of the information into a sequence of chronologic events (i.e., a timeline), facilitating a temporal visualization paradigm (Figure 1).

Accomplishing these tasks, however, requires sufficient lead time to retrieve patient data, organize the information, and initiate further data processing—in effect, prefetching is required. Alternative methods, described in other reports,[17–19] by which information is retrieved and integrated in real time on demand, do not solve issues pertaining to lengthy restoration times

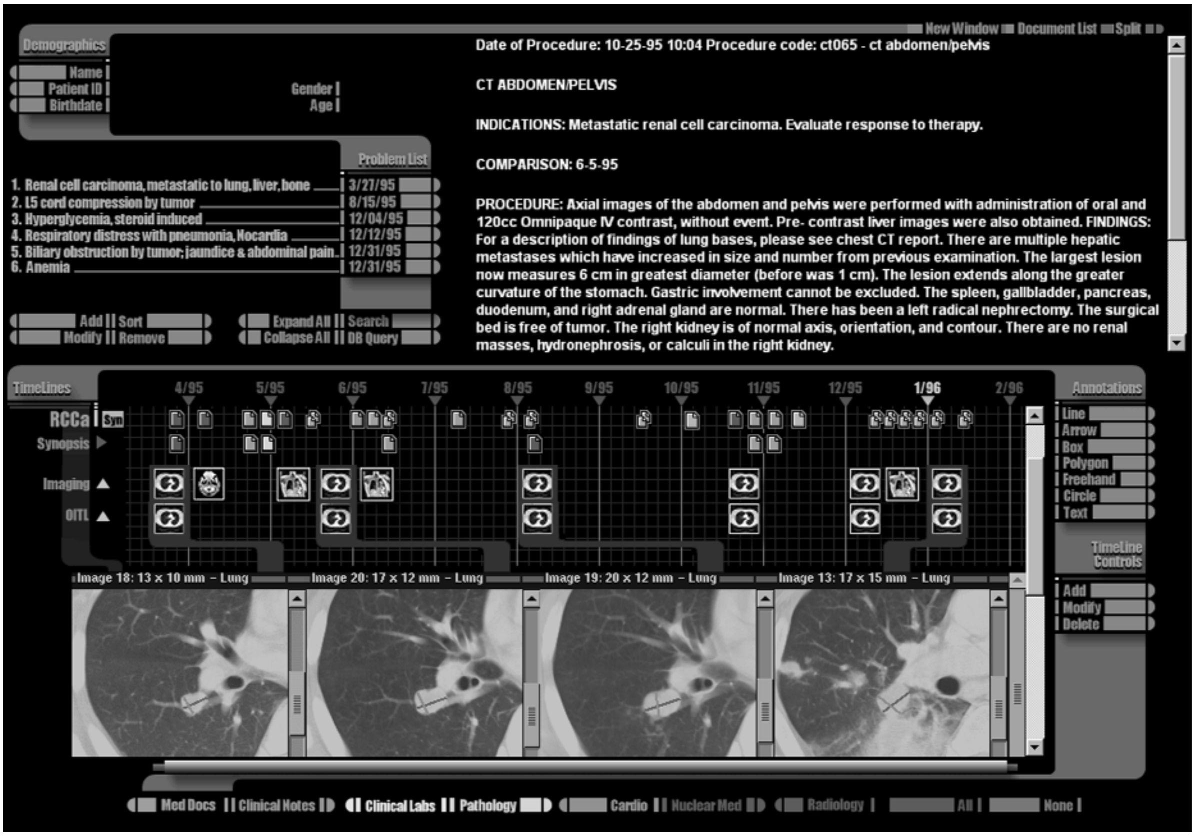

Figure 1 A screen of the Oncology Imaging Timeline (OITL) application supported through prefetching. OITL integrates PACS images with a patient's records from the hospital and radiology information systems to create a visual chronology of patient care. Tumors tracked by radiologists to determine radiographic response to treatment are recorded and displayed with other pertinent oncologic data.

of archived data (e.g., PACS) or the computational times required for the generation of derived data.

Conventional prefetching algorithms for imaging studies are predicated on information from scheduling and requisition data (e.g., anatomic region of interest, imaging modality, acquisition device, patient demographics), using this information to select the types of archived imaging to retrieve and determine where data should be routed.[20–23] For example, given an 8-year-old patient scheduled for a brain MR examination, rules encoded by a prefetching algorithm can control the routing of the image series and associated archived data to a pediatric radiology workstation; moreover, current prefetching techniques will restore only "related" imaging examinations (e.g., the interpretation of a brain MR examination will not likely require the review of previous chest CT examinations; thus, such examinations would not be restored from PACS). Indeed, current radiology workstations perform prefetching of historical patient data on the basis of simple rules, such as "Retrieve the most recent study with a given organ and/or imaging modality."[24,25]

To facilitate OITL's comprehensive view of patients' data, a new method was required to prefetch imaging and other clinical data on the basis of more abstract concepts not explicitly stored in PACS (e.g., "Patients with lung cancer managed by Dr. Jones").

OITL Prefetching

In this section, we detail the requirements that prefetching must fulfill, leading to specification of the developed architecture and its implementation.

Design Objectives

A prefetching mechanism, termed the prefetch monitor, was created to meet the requirements of OITL. The prefetch monitor performs the following activities.

Patient Identification. Patients whose data will be prefetched are selected by means of rules in a declarative language that allows high-level characterizations of patient cohorts (i.e., medical conditions, such as cancer). The developed rule language permits

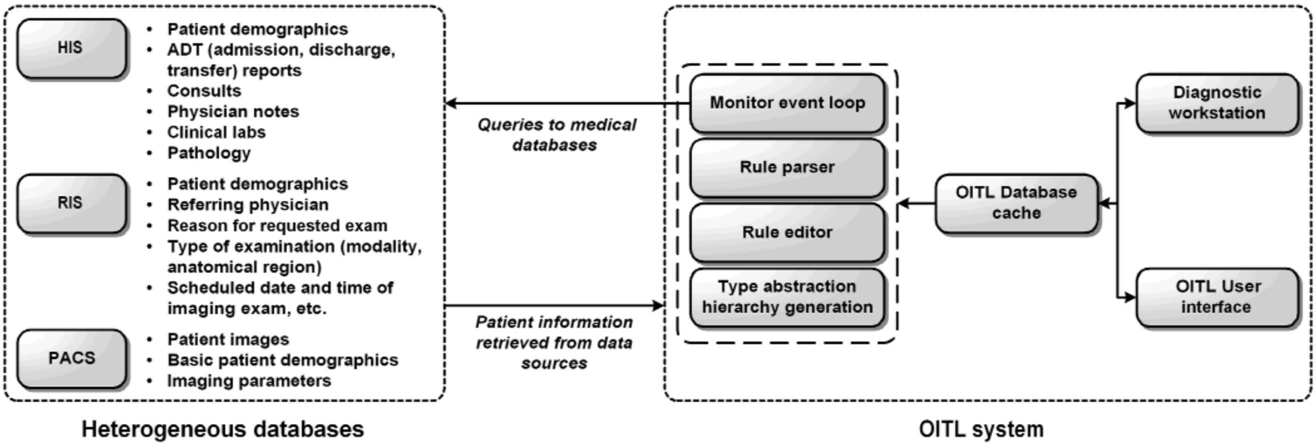

Figure 2 Components of the Oncology Imaging TimeLine (OITL) and the prefetch monitor. HIS indicates hospital information system; RIS, radiology information system; PACS, picture archiving and communication system.

patient identification using descriptions drawn from different fields stored across distributed medical databases. Rules can be created and customized by physicians through a simple interface. Electronic schedules are queried to detect new and add-on examinations to ensure availability of up-to-date patient data. Although efforts to promote codification and standardization of the contents of the medical record are ongoing, such endeavors are not complete nor fully adopted.*

Patient Data Retrieval. Once a patient has been identified, data are prefetched from all information repositories and integrated into a patient-specific timeline view.

Patient data are made accessible in advance (e.g., two weeks prior to the next scheduled clinic visit) to enable their review and to modify patient management considerations when appropriate. A limitation of traditional prefetching methods is that retrieval is typically triggered by (expected) patient arrival for an imaging examination or completion of a study; although this may ensure the availability of data coincident with radiographic diagnosis, case review may be desirable at times unrelated to these events. For example, an oncologist might want to review, with the radiologic consultant, data from a complex patient case some time before a scheduled visit—ideally, patient data would already be accessible at this time.

Most specialists work in designated clinics, where groups of computers are shared, typically on a local area network (LAN). In such an environment, it is sufficient to retrieve information to one of these machines (i.e., a prefetch monitor in each clinic), which effectively functions as a local server/cache, permitting end-users to retrieve prefetched patient data from clients in their clinical environment. In this paper, when we refer to prefetching and local caching to a workstation, we refer to prefetching to this server.

The end-users of OITL are medical and surgical oncologists as well as radiologists; prefetching is configured accordingly for each group.

Feature Extraction. Time-consuming processing tasks, like image segmentation, can be completed prior to physician review. For instance, automated knowledge-based segmentation of a chest CT can take, on average, more than 20 minutes[26]; having this image processing completed beforehand is advantageous.

Architecture and System Implementation

The components of OITL and the prefetch monitor are diagrammed in Figure 2. A prefetch monitor is located on each OITL server, permitting independent regulation of the data fetched to clusters of computers. The prefetch monitor is linked to a local OITL database that maintains prefetched information from the hospital and radiology information systems, images from the PACS, and derived data (image segmentation), all organized into a timeline view. The OITL database is then accessed by the OITL interface or the diagnostic workstation, or both.

Since data are cached on a local database server to which only authorized client programs (i.e., the OITL

*That is not to say that codes are not used in reporting. For example, ICD-9 codes sometimes appear in free-text reports to facilitate billing, but such codes are not stored in separate, searchable fields. Moreover, current manual coding of procedures for billing (e.g., CPT) can take several hours or days to complete, beccause of personnel limitations, so it is not helpful in prospective identification of patients.

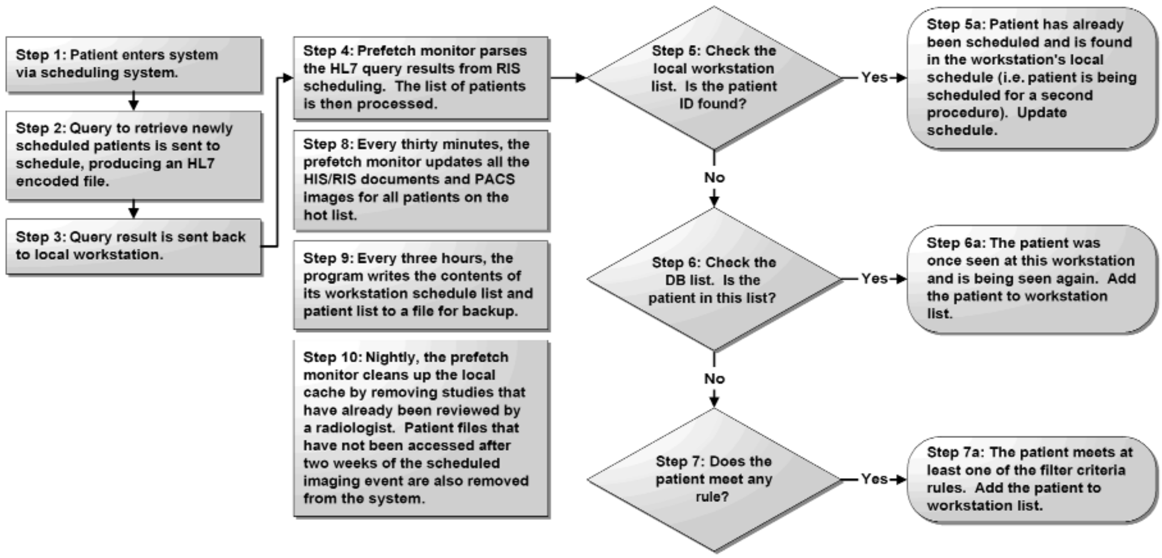

Figure 3 The patient detection cycle, which captures the basic algorithm used in finding prefetch candidates. HL7 indicates Health Level 7; RIS, radiology information system; HIS, hospital information system; PACS, picture archiving and communication system.

interface) can gain access, a level of security and confidentiality of patient records is maintained. This design permits multiple clients to utilize the prefetched data on the server, assuming a sufficiently fast network connection over which data can be transferred.

To perform its tasks, the prefetch monitor stores both retrospective and prospective lists of patients whose information has been, or will be, prefetched :

- *Local workstation schedule.* Since each workstation is customized for particular patient cohorts, the workstation has an independent list of patients whose information will be prefetched. This list is the entire prospective schedule of patients determined by the prefetch monitor; each patient name on this list has an associated date and time for a future examination.

- *Hot list.* The hot list is a subset of the local workstation schedule representing all patients who have examinations within two weeks before or after the current date. The hot list is used to determine which patient records should be queried to retrieve data for local caching, with continuous updating in anticipation of the clinician review process. Conversely, as patient names are removed from this hot list, the prefetch monitor removes the cached information from the workstation.

- *Database patient list.* The database patient list is a registry of all patients who were on the hot list and whose information was also subsequently reviewed by a physician at the workstation. As the names of some patients on a hot list may be false positives for a given patient cohort (i.e., patients who are thought to belong to a patient population but are not, in reality, a part of the group), the database patient list represents a continually growing true positive set (i.e., patients who were correctly identified as a part of a given cohort). Each patient whose name is on this list can be associated with one or more patient populations.

The following sections discuss in more detail the processes followed by the prefetch monitor.

Patient Detection Cycle

The core of the prefetching process is the patient detection cycle, which captures the basic algorithm used in finding prefetch candidates (Figure 3). The cycle starts with a query from the prefetch monitor to a scheduling system to retrieve information for newly scheduled patient examinations. Querying occurs at a predefined frequency (e.g., every 15 minutes) to ensure that information for patients added to the schedule at the last moment is incorporated into the prefetching process. The result of the query is an HL7-encoded patient list, which is then sent back to the OITL workstation's prefetch monitor (Figure 3, steps 1 through 4).

For scheduling image examinations from the radiology information system, for example, the information on the patient list consists of patient demographics (name, patient ID, etc.), procedural information (imaging modality, anatomic regions, scheduled date and time), referring physician, reason for requested procedure, and other related imaging examination information (e.g., contrast allergies). The types of queries sent by each prefetch monitor can be customized to the requirements of the workstation; for example, a review workstation with specialized software for analyzing CT studies may augment the base query by the additional constraint that only information for patients scheduled for CT examinations should be returned.

The patient information returned from the scheduling query is filtered to determine candidacy for prefetching. First, two heuristic searches are performed on the patient information returned from the scheduling query†:

- First, patient information is compared with the local workstation schedule (Figure 3, step 5). Finding a newly scheduled patient in the local workstation schedule implies that the patient was previously scheduled *and* that it was determined earlier by the system that the patient information should be prefetched; the resultant action is to add the patient information again to the local workstation schedule for the new event.
- Second, if the patient information is not found in the local workstation schedule, the database patient list is then searched (Figure 3, step 6), on the assumption that if patient information was previously prefetched and reviewed at the workstation, it should again be prefetched. Discovery of the patient information in the database patient list also causes the patient information to be added to the local workstation schedule.

Failing these two searches, a set of rules used to identify candidate patients are applied to the patient information (Figure 3, step 7). Physicians often use implicit information to identify members of patient cohorts (e.g., Patient Smith is referred by Dr. Jones, who specializes in thoracic oncology, and so Mr. Smith is possibly a lung cancer patient); this knowledge is expressed in the rule language developed for the prefetch monitor (see Filtering Using Rules, on this page). Patient information that fulfills at least one rule is added to the local workstation schedule.

Maintenance and Update Cycle

In parallel with the detection of newly scheduled patients, the maintenance and update cycle is responsible for retrieving the associated patient data and maintaining the workstation's local cache.

The prefetch monitor queries hospital databases every 30 minutes for any new, non-cached documents, laboratory reports, and other non-imaging data for patients whose information is on the local workstation schedule (Figure 3, step 8). PACS is queried to retrieve imaging data to the workstation for patients on the hot list. Also, updates to any prefetched data are determined (through examination of time stamps) and "invalidated" data are replaced to provide up-to-date information. As data are cached in the OITL database server, preprocessing tasks associated with the different types of data, such as image processing, are initiated as background processes. This part of the cycle helps guarantee that the most current data for a patient are available on the OITL workstation. As the data are retrieved from the different data sources, they are indexed chronologically to support the temporal visualization of the OITL user interface.

Every three hours, the prefetch monitor creates a backup copy of the current contents of its local workstation schedule (Figure 3, step 9); in the event that the prefetch monitor is disrupted, the system can resume from its last checkpoint and minimize disruption to the clinical operation of OITL.

Finally, the prefetch event loop performs maintenance nightly (Figure 3, step 10). As patients' names are removed from the hot list or as their data are reviewed, their data are assumed to be no longer required for review and are deleted from the local cache. Moreover, patients who were on the hot list and whose data were reviewed by a physician are added to the database patient list.

Filtering Using Rules

As stated before, the prefetch monitor uses a set of declarative rules to filter the results of a scheduling query. Each rule defines the characteristics of individual patients belonging to a targeted cohort by using data stored in multiple information sources (hospital information system, radiology information system, PACS). The conditions of each rule are examined in relation to each scheduled patient, searching databases indicated in the rule to determine whether a patient meets the given criteria. Patients whose information satisfies all conditions of a rule are added to the local workstation schedule. At a high level, for instance, a rule may instruct the prefetch monitor to search a patient's hospital information documents to find mention of chemotherapy treatments.

† These searches are deemed heuristic, since they assume that patient information that was seen once on the workstation is likely to be seen again and should be prefetched; this assumption is not always true but appears to be a sound guideline.

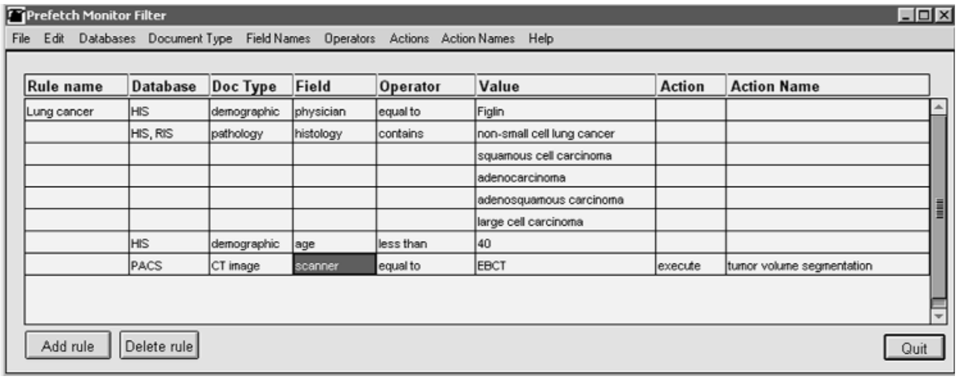

Figure 4 User interface designed for creating and editing the filter file for the prefetch monitor.

To facilitate the searching of data in a heterogeneous environment, a data directory is used, mapping components of the patient record to the corresponding data source. For example, the data directory contains information on where the prefetch monitor obtains patient demographics (hospital information system) vs. radiology reports (radiology information system). Unique field names are attributed to each entry in the data directory (e.g., the field name clinicalLab is mapped to the subset of the hospital information system pertaining to clinical laboratory tests and values). The data directory includes fields from the DICOM (Digital Imaging and Communications in Medicine) standard to support PACS. New types of information can be added to the prefetch monitor's search space by extending the data directory.

Rules take the following form:

RULE rulename IF conditions ACTION action

A rule starts with the keyword RULE and is followed by a rulename that can be used for referential purposes by users. The conditions of the rule are a Boolean conjunction of constraints, with each constraint stated as follows:

FIELDNAME operator value

FIELDNAME represents an established fieldname in the prefetch monitor's data directory. The fieldname is associated with a comparison operator and a value, allowing binary comparisons between the patient's data and the stated value. For fieldnames representing numeric values (e.g., age), standard comparators are supported (i.e., <, ≤ , =, ≥ , >, ≠). For text-based fieldnames, regular and substring (in)equality searches are permitted. An optional action can be specified for the rule using the ACTION keyword. The action is executed when the rule is triggered. A collection of predefined actions is provided with the prefetch monitor to initiate electronic pages, e-mail, and other alerts. To make the prefetch monitor extensible, the program's run-time library can be supplemented with new code implementing additional actions.

The following are examples of valid rules that can be established for the prefetch monitor:

- RULE rule1 IF (referringPhysician="Chan" AND clinicLocation | ``Laloma" AND patientAge > 25 AND patientGender="M")

  This rule. rule1, instructs the prefetch monitor to look for any patients whose referring physician is Dr. Chan, from clinics with the substring "Laloma" in the name (e.g., "St. Laloma Clinic", "Laloma North"), who are older than age 25 and male; all four conditions need to be true for the rule to be triggered. The prefetch monitor's data directory is intelligent enough to recognize that patient age is calculated from the patient birthdate; thus, the data directory contains instructions to find the birth date to establish the current age.

- RULE rule2 IF (patientConsults | "non small cell lung cancer" AND patientConsults ≠"adenocarcinoma")

  The second rule illustrates the use of a rule to search clinical consults stored in the medical record, for patients who have the keyword "non small cell lung cancer" but who do not have "adenocarcinoma" mentioned in any of the documents.

- RULE rule3 IF patientID="201-529-419" ACTION pageReferringDoctor

  This last rule, rule 3, shows the use of the optional ACTION clause. In this example, if a patient (defined by specific ID) is scheduled for an examination, the prefetch monitor will page the referring doctor.

Rather than creating permanent (i.e., "hard-coded") filters for the prefetch monitor, a set of rules is specified in a filter file. This filter file is quickly parsed each time results from the scheduling query are returned, allowing real-time updates of the prefetching system without interruption of its actual operation.

User Interface

Figure 4 shows a simple user interface designed for creating and editing the filter file for the prefetch monitor. Although the file is actually a plain-text file in syntax of the declarative rule language, the premise behind this graphical interface is to enable end-users (such as physicians or medical technicians) to customize the behavior of the workstation as to which patients are prefetched. End-users can readily add, edit, and delete rules from the system. Menus eliminate the need for a user to remember field names or operators in order to specify the conditions of a rule.

Patient Record Augmentation

The prefetch monitor effectively groups similar patients based on the satisfaction of rules. If a rule(s) describes a medical condition (e.g., lung cancer), it is reasoned that only patients who have the same characteristics will trigger the same rule. Using a data structure known as a type abstraction hierarchy,[27] it is possible to further cluster patients within a single patient population; each resulting "sub-cluster" is a collection of patients who have various similar user-defined attributes and measures.

The OITL database is based on the Knowledge-based Multimedia Medical Database (KMeD).[28] KMeD provides facilities for the approximate matching of complex objects using type abstraction hierarchies. The OITL prefetch monitor can generate type abstraction hierarchies using subsets of patients who trigger the same rule, resulting in definitions of subpopulations within a given patient cohort. Ultimately, as a physician accesses information on a given patient, it is possible to use the type abstraction hierarchies to provide supplemental information (e.g., patients under similar therapy) that may be useful in evaluating a particular patient or medical condition.

Evaluation

To test the effectiveness of our prefetching mechanism, an initial assessment of the system was conducted with two aims—to evaluate the accuracy of the prefetching algorithm, and to investigate whether using descriptors from multiple data sources improves the ability of prefetching to correctly identify members of a patient cohort.

Methods

Prefetching

Prefetching was focused on a target cohort from the oncology outpatient clinic. Specifically, the prefetch monitor was instructed to fetch the information for all lung cancer patients who were being seen by any one of three (of a possible five) oncologists at the clinic and who had a scheduled radiologic assessment. Since the majority of lung cancer patients treated at UCLA receive thoracic CT imaging to track response to treatment, the prefetch monitor's rules were configured to query scheduling in the radiology information system for patients scheduled for these procedures. The patient information returned in the query results was then filtered using the following conditions: 1) referring physician name matching one of the chosen oncologists, and 2) any mention of the keywords "lung cancer" or "lung carcinoma" in the patient's medical documents in the hospital information system.

Testing targeted a four-month period. However, since radiology procedures can be scheduled several months in advance of the actual study date, the five previous months were also examined to accurately capture all the scheduled examinations made during the evaluation period.

Effect of Multiple Data Sources on Prefetching

To test our hypothesis that using multiple data sources to describe individual patients provides better discrimination of patient cohorts, we examined the effect of limiting the prefetching rules to single data sources.

Results

Correctness of the prefetching methodology was determined by retrospectively comparing the oncology clinic's actual schedule against the resultant hot lists and database patient list generated by the prefetch monitor. In essence, validation of the prefetching results was judged by determining whether a prefetched patient was truly seen at the lung cancer clinic with an associated imaging procedure. Classic information retrieval measures were calculated to gauge prefetching performance, defined as follows:

$$\text{Recall} = \frac{\text{Total number of scheduled patients' records retrieved}}{\text{Total number of patients scheduled at the oncology clinic}}$$

$$\text{Precision} = \frac{\text{Total number of scheduled patients' records retrieved}}{\text{Total number of patients retrieved}}$$

Table 1 ■

Summary of Evaluation for Prefetching Performance

Description	No.
Number of CT exams processed by prefetch monitor during 9-month test period (5-month lead time plus 4-month targeted evaluation interval)	4,681
Number of patients after filtering radiology information scheduling system query for referring physician and target date	293
Number of patients remaining after additional filtering with keyword search and prefetched by the Oncology Imaging TimeLine (OITL)	40
Number of patients manually determined to be in target cohort	20
Number of patients that were both prefetched and in the target cohort	20
Number of patients prefetched that were not in the target cohort	20

Prefetching

A total of 4,681 thoracic CT examinations (first constraint) were performed during the evaluation period. Of the patients on whom these imaging procedures were performed, 293 were seen by one of the three specified oncologists (second constraint). The third constraint, which involved searching these patients' medical documents for keywords (lung cancer, lung carcinoma), produced a final list of 40 different patients whose information was prefetched to the OITL workstation during the testing period.

Within the targeted 4-month time interval, the oncology clinic's schedule revealed a total of 89 different lung cancer patients were seen by the three oncologists targeted by prefetching. Of these 89 patients, only 20 were scheduled for CT imaging studies; thus, these 20 patients formed the actual target cohort (i.e., truth).

The resulting comparison showed that for all 20 of the oncology clinic's patients with imaging, prefetching was correctly performed by the system,[†] indicating a recall of 100 percent for the prefetching algorithm. However, OITL incorrectly prefetched information for many patients who did not fall into this lung cancer cohort and thus had low precision (50 percent). The results of the analysis are summarized in Table 1.

† Several of these patients were, in actuality, prefetched multiple times because of sequential imaging to track progress; for each event, OITL correctly performed prefetching of the patient's data.

We anticipated that the number of false-positive results (i.e., incorrectly prefetched patients) would be high on the basis of the constraints of our prefetch rules. Particularly, the application of a keyword filter rule does not provide proper contextual searches. For example, searching for the keywords lung cancer or lung carcinoma returned both positive findings (e.g., "patient was diagnosed with lung cancer") and negative findings (e.g., "there is no family history of lung cancer"). We chose to err on the side of over-inclusion on the grounds that the assurance of timely data access justified the additional caching of non-targeted patients.

Multiple Data Sources

When we restricted the prefetching rules to information in the radiology information system only and performed the same evaluation on the oncology clinic schedule, the prefetch monitor retrieved only 6 of the 20 targeted patients—results that have perfect precision (i.e., no extraneous patients were prefetched) but a very low recall (30 percent). In comparison, when the prefetch monitor's rules were restricted to information solely from the hospital information system, only 13 of the 20 lung cancer patients were prefetched (65 percent recall). Thus, in both scenarios, neither database (the radiology or the hospital information system) provided sufficient information to properly identify candidate patients, and some patients were not prefetched.

This result may be explained by pointing out that a patient cohort may be sufficiently defined only by combining data from different databases (e.g., imaging procedure type from the radiology information system and referring physician from the hospital information system); independently, neither database can provide sufficient constraints to identify members of a patient population effectively. This observation, we believe, underscores the benefit of our multiple data source approach to prefetching.

Although this evaluation showed that the prefetching rules provided high accuracy, it was somewhat simplistic in the idea that the diagnoses of targeted patients could be captured by specific keywords. One possible problem with this rule-based methodology is the specification of cohorts with a broader diagnosis category (e.g., "oncology") for which numerous keywords might be required. We believe that this issue can be resolved by linking keywords to an ontology or standardized terminology (e.g., the UMLS or SNOMED), whereby synonyms and related terms can be automatically incorporated into prefetch monitor searches; this idea is currently being investigated by use of natural language processing techniques.[29,30]

Discussion

We have described a problem-oriented prefetching system to facilitate patient management and review by physicians. While examples have been drawn mainly from the oncology domain, the system can be adapted to other medical domains (e.g., orthopedic, urology). Prefetching of imaging data is supplemented by retrieval of patient data from other hospital databases; the totality of the information is integrated into a comprehensive overview and made available to physicians. The prefetching method emphasizes the accuracy of patient selection to ensure availability of data for all patients prior to review by managing clinicians.

Previous and Related Work

Prior work on prefetching varies from manual methods requiring human intervention[31] to fully automated mechanisms. We briefly review three other approaches described in the literature and provide a comparison of these efforts with the OITL prefetching system.

Second-generation Picture Archiving and Communication System at the University of California, San Diego

The University of California, San Francisco (UCSF) second-generation PACS was a system designed to optimize image retrieval, archiving, and clinical operations.[20,32–34] A component of this PACS was a prefetching system using admission, discharge, and transfer messages from the hospital information system as a triggering mechanism; that is, as patients were admitted to the hospital, the hospital information system transmitted a message to the radiology information system, prompting the insertion of patient demographic data into the radiology information system and initiating prefetching. Using a lookup table based on imaging examination type, disease category, referring physician, and radiologist, historical PACS studies were recalled from long-term optical storage. The number and age of past imaging examinations influenced which studies were retrieved for review by the radiologist. Additional parameters in the table guided the prefetching of other radiology information system data (e.g., radiology reports). This work was more recently extended to address better determination of relevant priors.[35]

Image Retrieval Expert System

The Image Retrieval Expert System (IRES) project uses a knowledge base to determine what images are pertinent for review, using two sets of factors: 1) imaging examination parameters, including the study type and the associated reason for examination (i.e., reason for request); and 2) the preferences of the individual user (i.e., radiologist).[21,36] The IRES knowledge-base stores classifications of reasons for imaging studies, anatomic procedure information (e.g., abdominal CT), and temporal relationships that exist between previous imaging series (e.g., the last abdominal CT examination was three years prior). Heuristic (domain) information stored in the knowledge base is determined using a neural net trained to observe individual radiologists' image retrieval patterns in order to later predict what type of images should be prefetched under similar circumstances.

Picture Archiving and Communication System at the University of California, Los Angeles

Work on the UCLA PACS has established a prefetching system that uses combinations of admission, discharge, and transfer messages from the hospital information system and event registration messages from the radiology information system (e.g., examination scheduling, patient arrival, completion of imaging procedures) to trigger retrieval of archived images.[22,37] The UCLA PACS is based on a cluster architecture, in which each cluster is logically defined as a group of computers. Each cluster has a cluster controller, responsible for the routing and prefetching of images. The set of images retrieved from storage is determined by comparing the "correlation" of the current imaging examination with past examinations. The correlation value is based on the similarity between a past examination's characteristics (date of study, associated organ system, imaging parameters, referring physician, rendering radiologist) and the current imaging study.

Comparison of Prefetching Methods

The OITL prefetching approach differs from earlier algorithms in several ways:

- *Tailored patient identification.* Specific patient populations can be described using high-level abstractions, such as medical problems, utilizing information from different data sources (e.g., hospital information system, radiology information system, PACS). The OITL prefetching rules are configurable according to different abstractions, whether retrieval is patient-oriented, problem-oriented, or based on any combination of desired data attributes. This flexibility in defining rules can be used in supporting the development of clinical testbeds for outcomes or other health services research.

In contrast, the UCSF, IRES, and UCLA PACS prefetch mechanisms do not support the abstract description of patients using multiple data sources; as such, unless medical conditions are codified in the patient record, automated prefetching for a given medical condition (i.e., cohort) is not possible.

- *Multimedia prefetching.* Conventional prefetching is principally concerned with retrieval of archived imaging, with less consideration given to retrieval of other patient data (e.g., hospital documents, laboratory reports) contained in medical information systems. OITL performs prefetching on all patient-related data.
- *Timing and updating of patient data.* OITL prefetching enables the system to present up-to-date data as soon as a patient is scheduled in the system, as opposed to when the patient physically comes to the hospital or radiology department for a procedure. Earlier prefetching of the data facilitates review of patient information prior to the actual patient encounter.

Conclusion

Prefetching appropriate data according to a patient's medical condition allows convenient review of the integrated information by the clinician.

We have designed and implemented the OITL prefetch monitor, changing the focus of prefetching from a patient-oriented to a problem-oriented mechanism. The prefetch monitor is capable of retrieving hospital information system, radiology information system, and PACS data for a described patient population, locally caching imaging and reports well in advance of anticipated need. Initial testing of the system has shown high recall but low precision. Evaluation also revealed that the use of multiple data sources to describe patients and aid in their categorization into patient cohorts improves prefetching performance. Our approach can be utilized to collect and integrate information for any clinic where patients are routinely scheduled.

Future work on the system will include extending the rule language to handle regular search string patterns, enabling search capabilities such as those found in well-known Unix programs like awk and grep. We are also exploring the use of natural language processing for those medical domains with established lexicons. The use of natural language processing on free-text reports to handle the context of findings may increase the level of precision in document searching and facilitate searching of related terms for diagnoses. Finally, since we have not yet addressed the prefetching of data for review from non-traditional clinical environments (e.g., a physician's home), we are adapting this prefetching technique to Web servers, permitting smaller amounts of data (lab reports, documents, selected PACS images) to be cached and more quickly accessed via a browser.

References ■

1. Kangarloo H. PACS: clinical experience at UCLA. Comput Med Imaging Graphics. 1991;15(3):201–3.
2. Huang HK, Kangarloo H, Cho PS, Taira RJ, Ho BK, Chan KJ. Planning a totally digital radiology department. AJR Am J Roentol. 1990;154(3):635–9.
3. Eversman WG, Pavlicek W, Zavalkovskiy B, Erickson BJ. Performance and function of a desktop viewer at the Mayo Clinic Scottsdale. J Digital Imag. 2000;13(2):147–52.
4. Ratib O. From multimodality digital imaging to multimedia patient record. Comput Med Imaging Graphics. 1994;18(2):59–65.
5. Forslund DW, Phillips RL, Kilman DG, Cook JL. Experiences with a distributed virtual patient record system. Proc AMIA Annu Fall Symp. 1996:483–7.
6. Kilsdkonk ACM, van der Weriff A. Structuring the patient record: NUCLEUS (customization environment for multimedia integrated patient records). Comput Methods Prog Biomed. 1994;45:127–30.
7. Santoni R, Cellai E. Multimedia clinical records: results of a pilot project at the Radiation Therapy Department of Florence. Br J Radiol. 1995;68:413–20.
8. Lowe HJ, Buchanan BG, Cooper GF, Vries JK. Building a medical multimedia database system to integrate clinical information: an application of high-performance computing and communication technology. Bull MLA. 1995;83(1):57–64.
9. Lowe HJ, Buchanan BG, Cooper GF, Kaplan B, Vries JK. Image engine: an integrated multimedia clinical information system. MedInfo. 1995;8(pt 1):421–5.
10. Aberle DR, Dionisio JDN, McNitt-Gray MF, et al. Integrated multimedia timeline of medical images and data for thoracic oncology patients. Radiographics. 1996;16:669–81.
11. Bui AAT, Aberle DR, McNitt-Gray MF, Cardenas AF, Goldin JG. The evolution of an integrated timeline for oncology patient healthcare. Proc AMIA Annu Symp 1998:165–9.
12. Cousins SB, Kahn MG. The visual display of temporal information [technical report]. St. Louis, Mo: Washington University School of Medicine, Medical Informatics Laboratory, 1991.
13. Plaisant C, Milash B, Rose A, Widoff S, Shneiderman B. Lifelines: visualizing personal histories. Proc ACMCHI Conference. 1996:221–7.
14. Zhu X, Lee K-N, Levin DL, et al. Temporal image database design for outcome analysis of lung nodules. Comput Med

Imaging Graphics. 1996;20:347–56.
15. Dionisio JDN, Cardenas AF. A unified multimedia data model for representing multimedia, timelines, and simulation data. IEEE Trans Knowledge Data Eng. 1998;10(5):746–67.
16. Dionisio JDN, Cardenas AF, Taira RK, et al. A unified timeline model and user interface for multimedia medical databases. Comput Med Imaging Graphics. 1996;20(4):333–46.
17. Clinical Multimedia Laboratory Center for Biomedical Informatics. Image Engine: system architecture > agents. 2000. University of Pittsburgh Medical Center Web site. Available at http://www.cml.upmc.edu/cml/ImageEngine/agents.html. Last accessed Mar 14, 2001.
18. Dayhoff RE, Kuzak PM. Providing complete multimedia patient data to consulting radiologists, other specialists, and the referring clinician. J Digital Imaging. 1998;11(3):134–6.
19. Dayhoff RE, Kuzak PM. Kirin G, Frank S. Providing a complete online multimedia patient record. Proc AMIA Annu Symp. 1999:241–5.
20. Wong AWK, Huang HK, Arenson RL, Lee JK. Digital archive system for radiologic images. Radiographics. 1994; 14(5):1119–26.
21. Sheng ORL, Wei C-P, Ozeki T, Ovitt T, Ishida J. Design of knowledge-based image retrieval system: implications from radiologists' cognitive process. Proc SPIE Conference on Medical Imaging: PACS Design and Evaluation. 1992:243–54.
22. Chao W, Ho BKT, Chao JT, Sadri R, Huang L, Taira R. Implementation of system intelligence in a 3-tier telemedicine/PACS hierarchical storage management system. Proc SPIE Conference on Medical Imaging: PACS Design and Evaluation. 1995:423–9.
23. Wilson DL, Smith D, Rice B. Intelligent prefetch strategies for historical images in a large PACS. Proc SPIE Conference on Medical Imaging: PACS Design and Evaluation1994:112–23.
24. Ho B, Chao W, Sadri R, Huang L, Taira R, Shuh H. Data clustering and other archival retrieval strategies for teleradiology and picture archiving and communication systems. J Digital Imaging. 1995;8(4):180–90.
25. Amicas Web/Intranet Image Server: teleradiology brief. Amicas Inc. Web site. 1999. Formerly available at http://www.amicas.com/support. Last accessed 1/99.
26. Brown MS, McNitt-Gray MF, Mankovich N, et al. Method for segmenting chest CT image data using an anatomical model: preliminary results. IEEE Trans Med Imaging. 1997;16(6):828–39.
27. Chu WW, Chiang K, Hsu C-C, Yau H. An error-based conceptual clustering method for providing approximate query answers. Commun ACM Virtual Extension Ed. 1996.
28. Chu WW, Cardenas AF, Taira RK. KMeD: a knowledge-based multimedia medical distributed database system. Inf Syst. 1995;20:75–96.
29. Taira RK, Soderland SG. A statistical natural language processor for medical reports. Proc AMIA Annu Symp. 1999:970–4.
30. Sinha U, Yaghmai A, Dai B, et al. Evaluation of SNOMED 3.5 in representing concepts in chest radiology reports: integration of a SNOMED mapper with a radiology reporting workstation. Proc AMIA Annu Symp. 2000:799–803.
31. Henri CJ, Cox RD, Bret PM. Implementation of a filmless mini picture archiving and communication system is ultrasonography: experience after one year of use. J Digital Imaging. 1997;10(3 suppl 1):80–2.
32. Wong AWK, Huang HK, Arenson RL, Yin L. Automated prefetch mechanism design and implementation for radiology PACS. In Proc SPIE Conference on Medical Imaging: PACS Design and Evaluation. 1994:103–111.
33. Wong AWK, Huang HK, Arenson RL. Hierarchical image storage management: design and implementation for a distributed PACS. Proc SPIE Conference on Medical Imaging: PACS Design and Evaluation. 1995:61–6.
34. Huang HK, Arenson RL, Lou ASL, et al. Second-generation picture archiving and communication systems. Proc SPIE Conference on Medical Imaging: PACS Design and Evaluation. 1994:527–35.
35. Andriole KP, Avrin DE, Yin L, Gould RG, Luth DM, Arenson RL. Relevant priors prefetching algorithm performance for a picture archiving and communication system. J Digital Imaging. 2000;13(2):73–5.
36. Sheng ORL, Wei C-P, Hu PJ-H, Sutjahjo SA, Han T-D. Knowledge-based image retrieval: a new generation design. Proc SPIE Conference on Medical Imaging: PACS Design and Evaluation. 1994:137–47.
37. Taira RK, Wong AWK, Stewart BK, Huang HK. Design of a PACS cluster controller. Proc SPIE Conference on Medical Imaging: PACS Design and Evaluation. 1992:203–7.

Problem-oriented Prefetching for an Integrated Clinical Imaging Workstation

Alex A T Bui, Michael F McNitt-Gray, Jonathan G Goldin, et al.

J Am Med Inform Assoc 2001 8: 242-253
doi: 10.1136/jamia.2001.0080242