High Speed Clinical Data Retrieval System with Event Time Sequence Feature

著者	Kimura Michio, Tani S, Watanabe H, Naito Y, Sakusabe T, Watanabe H, Nakaya J, Sasaki F, Numano T, Furuta T, Furuta T
journal or publication title	Methods of Information in Medicine
volume	47
number	6
page range	560-568
year	2008-11-20
URL	http://hdl.handle.net/10271/2333

doi: 10.3414/ME9125

High Speed Clinical Data Retrieval System with Event Time Sequence Feature

With 10 Years of Clinical Data of Hamamatsu University Hospital CPOE

M. Kimura¹, S. Tani¹, H. Watanabe¹, Y. Naito¹, T. Sakusabe², H. Watanabe³, J. Nakaya⁴, F. Sasaki⁵, T. Numano⁶, T. Furuta⁶, T. Furuta¹

¹Hamamatsu University, School of Medicine, Hamamatsu, Japan

²Fujita Health University, School of Health Sciences, Toyoake, Japan

³The University of Tokyo, Dept. of Planning, Information and Management, Tokyo, Japan

⁴Tokyo Medical and Dental University, Information Center for Medical Sciences, Tokyo, Japan

⁵NEC Corporation, Tokyo, Japan

⁶NTT Data Tokai Corporation, Nagoya, Japan

Summary

Objectives: This paper illustrates a high speed clinical data retrieving system, from 10 years of data of operating hospital information system for the purposes of research, evidence creation, patient safety, etc., even incorporating time sequence of causal relations.

Methods: Total of 73,709,298 records of 10 years at Hamamatsu University Hospital (as of June 2008) are sent from HIS to retrieval system in HL7 v2.5 format. Hierarchical variable length database is used to install them.

Results: A search for "listing patients who were prescribed Pravastatin (Mevalotin and generic drugs, any titer)" took 1.92 seconds. "Pravastatin (any) prescribed and recorded AST > 150 within two weeks" took 112.22 seconds. Searching conditions can be set to be more complex, connected by Boolean operator and/or. This system called D*D is in operation at Hamamatsu University Hospital since August 2002. It is used for 48,518 times (monthly average of 703 searches). Neither searching, nor background export of data from HIS caused delay of routine operating CPOE.

Conclusions: Search database outside of routine operating CPOE, with daily export of order data in HL7 v2.5 format, is proved to provide excellent search environment without causing trouble. Hierarchical representation gives high-speed search response, especially with time sequence of events.

Keywords

Medical order entry systems, hospital information systems, evidence-based medicine, standards, clinical research

Methods Inf Med 2008; 47: 560–568

doi:10.3414/ME9125

1. Introduction

Hospital information systems have many kinds of precious data, i.e. examination results, prescriptions, diagnosis classifications, as well as patient's demographic information. Making full use of them gives us not only speedy and user-friendly daily practices, but also analytic results of clinical data, even appropriate alerts for patient's safety. On the other hand, however, this data is huge in volume, and is loaded with a heavily used routine operating system.

In this paper, an analytical clinical information retrieval system called D*D is described, which can provide high-speed clinical information retrieval and has already been in use for more than seven years. D*D has a unique structure of data housing system within itself, which enables high-speed retrieval. The structure has a time-sequence feature between facts to be retrieved, which makes causal-relation search so fast. Using these functions, D*D provides not only clinical evidence data, but also background searching of noteworthy prescription orders, or even the pairing of prescription and lab examination results, which work to contribute to a patient's safety.

2. To Pursue Higher Patient's Safety and More Clinical Evidence.

Hospital information systems started providing examination results, past prescriptions, as well as patient's demographic information to users of healthcare providers [1-5]. Computerized physician's order entry (CPOE) enabled physicians to enter many kinds of orders directly, which helps decrease medical errors [6-13].

The basic way in which CPOE contributes to enhance patient's safety is that physicians select drugs among the available list displayed, instead of handwriting which can cause misreading. As CPOE has lab results and diagnosed disease names, it would be helpful to check it at the time of admission. For example, cancer chemotherapy would be with checking leukocyte count at all times. This is made possible by high speed checking of both prescription orders and lab results.

This function also enables analytical data retrieval for creating clinical evidence. To produce clinical evidence, many hypotheses must be checked. For example, suppose a physician thought a certain anti-hyperlipidemia drug may cause hepatic side effects, for any reason such as pharmacokinetics. The physician would check CPOE data for patients who have been prescribed the drug. The data to check would be AST,

LDP, Bil, etc. or even their combination. High speed retrieval, no matter how large the archived data are, should be highly appreciated, as it allows users to check as many hypotheses as possible in a limited time. Furthermore, hypotheses require causal relations. The data to be checked should be “after medication, within a certain period”. CPOE has, of course, time stamps for every order, and results.

To summarize, the safety enhancing, and evidence creating clinical database should 1) allow multiple combinations of retrieval, 2) accommodate time sequence of facts, 3) provide high speed retrieval, and 4) make as much previous data available as possible, without interfering the fast response of operating CPOE.

3. Co-existence of Routine Practice and Reference Analytical Tests

In Japan, installation of CPOE started in the 1980s. The prevalence of CPOE in Japan is very high, about 90% in large hospitals (400+ beds) in 2006 [14-16]. Secondary use of the data of CPOE, i.e. prescriptions, lab results, etc. have been tried from the beginning [17-19]. As CPOE requires a fast response not to keep physicians waiting, it is often the case that most CPOE have only a few years of past data available for real time use, though most hospitals actually have years of past data in archives.

CPOE uses many kinds of tables to make valid order messages. For example for prescription orders, CPOE uses available drug table, noting their attributes with special caution, and order doctors' table with their qualification to prescribe certain drugs, etc. This type of use is suitable for table-based relational database system.

On the other hand, retrospective analysis of clinical data available from CPOE and EMR (electronic medical records) are rather limited in variety. Drug name, dose, timing, route, lab-test name, results, unit, standard value, diagnosis name, code, code table, version, etc. Signs and symptoms can be rich in variety. Although they are limited in variety,

Fig. 1 Hierarchical representation of drugs. Drug code used here, which has hierarchical semantics within the code, abstract matching can be easily done, i.e., search by top 7 digit 2189010 means pravastatins, including Mevalotin and other generic drugs.

they should be represented with structure, especially in a hierarchical structure [20, 21]. For example, when a user searches for anti-hyperlipidemia drug pravastatin sodium, Mevalotin®, the user likes to have both pravastatin sodium 5 mg and 10 mg. It is tiresome for the user to check for every titer

of the drug and input them all. Figure 1 shows an example of hierarchical representation of this drug. If the database allows higher-class object “pravastatin” as a search key, it cuts down on the input frequency for other same content pravastatin generic drugs. In Figure 1, semantic hierarchical drug code sys-

Fig. 2 Data from CPOE to D*D retrieval through gateway prescription, injection, lab results, disease diagnoses, and patient demographics data in HL7 v2.5 format

Fig. 3 "Any pravastatin sodium, 5 mg or 10 mg, prescribed patient". Search result: 2920 patients, search time: 1.92 seconds

tem is also shown, which is used in this system.

This hierarchical representation is also suitable for describing signs and symptoms in progress notes of medical charts, because the basic concepts are hierarchically described as taxonomy and it is natural for users to describe according to hierarchy.

The result of this is that secondary use of analytical database should be separated from the real time practice of CPOE. The operating system of CPOE is suitable for relational database with many tables, while secondary database should have the hierarchical data model, with variable length of instances.

4. Time Sequence Featured Clinical Data Retrieval System with 10 Years CPOE Data of Hamamatsu University Hospital CPOE

Hamamatsu University Hospital [22, 23] has 620 beds, and its average outpatient number is 1500. CPOE has been in operation since 1989. It is now fourth CPOE. In recent replacement/upgrading of CPOE, the export of patient's demographics, prescription and injection orders, lab results,

disease diagnoses in HL7 [24-26] format has continued. Now there is 10 years of such data.

For this accumulated data, the authors constructed a high-speed clinical data retrieval system with time sequence feature, called D*D (database of structured records and diagnoses).

Because of the reasons mentioned in preceding chapter, D*D is independently established from CPOE, importing HL7 format data from CPOE. Figure 2 shows the data flow between CPOE and D*D.

Fig. 4 "Hypercholesterolemia patients showing Total Cholesterol >270 and treated with pravastatin sodium and atorvastatin calcium (Lipitor) prescriptions, any titer." Click on "AND" turn the operator to "OR" which can make logical combinations for search.

4.1 Example of Retrieval, Basic Search

Figure 3 is a screen shot of D*D. The left top window shows the searching criteria. This case is for "any pravastatin sodium 5 mg or 10 mg prescribed patient". Push "search button" and then lower-left list appears, with result of 2920 patients in 1.92 seconds and lists the patient name (anonymized) which is a hyperlink to the patient's folio. Figure 4 is for a search "Hypercholesterolemia patients showing Total Cholesterol >270 and treated with pravastatin sodium and atorvastatin calcium (Lipitor) prescriptions, any titer". User can add more searching criteria among other prescriptions, lab results, diagnoses, combined with AND/

OR Boolean operation. This search took 15.08 seconds with 206 patients. A table and graphic form of Total Cholesterol results of the selected patient is shown in Figure 5.

4.2 Example of Retrieval, Time Sequence Feature

The last example of the basic search mentioned above actually means "Patients diagnosed with hypercholesterolemia at any time, being prescribed pravastatin sodium and atorvastatin calcium more than once and having recorded T-Chol >270 at any time". An anti-hyperlipidemia medicine pravastatin sodium is reported to cause a hepatic side-effect [27-29]. As a drug-induced side-effect monitoring of the patients

after the side-effect alert, and for finding the hypothesis with clinical evidence, what we like to have is, i.e., a list of "patients who have shown AST >150, within two weeks after pravastatin sodium medication". D*D is equipped with this time sequence feature.

Figure 6 shows the condition set as "pravastatin (or others, any titer), and AST >150, at any time", results of 277 patients in 18.48 seconds. Figure 7 shows the condition set as "pravastatin (or others, any titer) and AST >150 subsequently within 2 weeks", results of 83 patients in 112.22 seconds. On the right, one of the patient's AST in the graph, which shows sharp elevation of AST at 1997/11/04 after medication of 10/22 and 10/29. The conditions can be set in combination with Boolean operations, and the sequence time can be set to other intervals.

Fig. 5 Result of search at Figure 4, with chronographic table and graph of Total Cholesterol of selected patient

5. Specifications and Data Representation

As D*D in Hamamatsu University Hospital has received 10 years of lab results, prescription/injection orders, disease diagnoses, and patient's demographics in HL7 message from the hospital information system is shown in Figure 2, comprising 385,815 patient demographics (PID of HL7 message), 1,705,853 disease diagnoses (DIS), 64,725,104 lab results (OBX), and 6,892,526 prescription and injection orders (RXE), as of June 2008.

Database system environments are as follows: database: cache ver 5.0; CPU: Intel Xeon (3.2 GHz) CPU×2; main memory of 3 GB with database cash 1280 MB.

Within this, the above-mentioned ten years of data are stored in 35 GB of HDD. Cache is a progressed database of MUMPS which is hierarchical variable length data structure, and has many clinical applications [30-32]. Additional to its quickness of search, MUMPS is highly data storage space-efficient. Hamamatsu University Hospital CPOE uses 14 GB for one year of order data storage, compared to 35 GB for ten years in D*D.

Table 1 shows average monthly searches by users of Hamamatsu University. It is now in the eighth year of operation. Physicians use for clinical evidence, with surprise for quickness of the search, as they knew that causal relation search usually takes hours by other so-called data warehouse. Already, the users of this system found some new clinical

evidence. The pharmacy department uses it mainly for medication side-effect monitoring. As D*D has background daily search macro function, notably a side-effect check is done automatically every midnight, and pharmacy department checks only in the morning.

Within cache database, all records are represented in a hierarchical index. D*D has indices for 1) disease diagnoses, 2) lab results, 3) prescription/injection orders, and 4) patient demographics. Figure 9 shows some index trees. A search of "pravastatin sodium prescription and AST > 150" evokes both "pravastatin sodium prescribed patient listing" and "AST > 150 patient listing", and they are merged according to the Boolean operator (AND, this case) between them. Because of this hierarchical representation,

Fig. 6 "Pravastatin (or others, any titer), and AST > 150, at any time". Search results: 277 patients, search time: 18.48 seconds

search keys should be in higher ranks of the hierarchy, in order to make searching quicker.

At the same time, D*D has another hierarchical index for each of above information. Figure 10 is the hierarchy for time sequence search. Patient IDs and date are set in higher position than regular hierarchy in Figure 9. There are two reasons to have this redundant information. First is to show selected patient's information, as in the right-hand side of Figure 7. Second is to accelerate time sequence search. For "A, then within two weeks, B" search, first a regular search for A is evoked. Then with the patient's ID list of A result, D*D looks for the other hierarchy for B. As patient's ID and data are in very high ranks, system can elim-

inate unnecessary lab results checking for non-A patients. Table 2 shows the searching time for "pravastatin sodium and AST > 150 within 2 weeks" for two cases. One is using only the original hierarchy twice. The other is using both. Searching time is about 1/6 (64.83/405.88).

6. Discussions

In [11], which explores possible impacts which CPOE brings, main emphasis is on patient safety. Clinical database is more or less than secondary use of the data, as also shown in [15]. Through long years of use of this database in Hamamatsu University

Hospital, it is shown that clinical database does not interfere daily use of CPOE, and that it can serve also for the basis of patient safety.

Simultaneous application of a conventional tree structure index and an index of patient's ID and dates at higher ranks reduced search period in time series to 1/6. However, this idea was realized under the condition of massive data to a few search key items such as test results, prescriptions and injections, and the name of the disease. The preparation of an optimized structural index presuming all cases for a number of search key items under indefinite retrieval sequence was difficult. In other words, a database of this type can offer a very high-speed retrieval for a

Fig. 7 "Pravastatin (or others, any titer) and AST > 150 subsequently within 2 weeks". Search result: 83 patients, search time 122.22 seconds. Patient list shows a selected patient has "HIT" prescription twice, 1997/10/22, and 10/29, and graph of AST shows peak high value recorded on 1997/11/04, within two weeks of the medication

Fig. 8 Hierarchical representation of lab results and medications: for lab results – lab test code – date – results – patient ID – order number. For medications – drug code – date – patient ID – order number

usage expecting the search items and procedure, but, for example, it seems to be inappropriate for an application to operate checking numerous types of information such as attributed classes of doctors (specialists or not), relationship between prescriptions and name of diseases (relations to indications).

The authors provided retrievals focused on the names of diseases, prescriptions and injections, and test results capable of being input with HL7 v2.5 without the consideration of medical charts such as findings by doctors and medical service personnel. This is because there was no establishment of a standard format or data type for taking in such information. While standardization of finding items in electronic medical charts such as HL7 CCD (Continuity Care Document) [33, 34] is advanced, its further development is awaited.

7. Conclusion and Future Plans

A database useful for creating drug safety, clinical studies and evidence has been accomplished through prescriptions and injections, test results, name of disease being manifested in a hierarchical structure. Operations in Hamamatsu University Hospital for seven years showed a smooth increase of users (in 2007, we had system replacement, which caused drop of use). In addition, no drop in response of the operating system occurred by occupying a network for data delivery from the hospital information system.

Time series retrieval, such as AST > 150, within two weeks of administration of pravastatin sodium, (Mevalotin®), made indications of causal relationships of events possible. Particularly, a combination with an optimized different tree-structure data realized quick retrieval. The quick retrieval gives users the chance to test various hypotheses. In other words, an example of hepatic disorder shown above gave users the opportunity to test such as ALT, LDH and TP for examinations but could not find a causal relationship in AST examination.

Other lessons learned are: 1) Cache database is compact, compared to other relational database; 2) index sequence in hierarchical structure is important to make search quicker.

A check for the basis of drug safety information has already been carried out, and development of a function capable of automatically informing by e-mail about any abnormal test results to treat situations such as a breakout of a new infection in the future is planned.

Also, it is now the case that ten years of data has been made retrievable, so compact even with a notebook computer, whereas until now it had only been realized through a server in a large computer system.

Acknowledgment

Accumulation of the vast amount of clinical information is owed to efforts of all staff members of Hamamatsu University School of Medicine Hospital and I wish to express my deepest appreciation.

Table 1 Monthly average and total search use of this system at Hamamatsu Hospital

2002 (Aug -)	2003	2004	2005	2006	2007	2008 (- Apr)	Total
454.8	627.3	775.5	712.8	886.8	649.8	604.0	48,518

Fig. 9 Hierarchical representation of lab results for time sequence search – patient ID – date – lab test code – result – order number

Table 2 Search time (ms) with/without second hierarchy. Test condition: data: 55,587,217 records (March 2007) (about 75% of today); CPU: Intel Core Duo T2300 (1.66 GHz); main memory: 1 GB with database cache 256 MB

	1st	2nd	3rd	4th	5th	Average
Without second hierarchy	405.55	405.89	406.75	406.03	405.17	405.88
With second hierarchy (Fig. 9)	65.42	65.11	65.00	64.20	64.42	64.83

References

1. Bakker AR. The development of an integrated and co-operative hospital information system. *Med Inform (Lond)* 1984; 9 (2): 135-142.
2. Safran C, Slack WV, Bleich HL. Role of computing in patient care in two hospitals. *MD Comput* 1989; 6 (3): 141-148.
3. McDonald CJ, Tierney WM, Overhage JM, Markin DK, Wilson GA. The Regenstrief Medical Record System: 20 years of experience in hospitals, clinics, and neighbourhood health centers. *MD Comput* 1992; 9: 206-217.
4. Rector AL, Nowlan WA, Kay S. Foundations for an Electronic Medical Record. *Methods Inf Med* 1991; 30: 179-186.
5. Ohe K, Kaihara S. Implementation of HL7 to client-server hospital information system (HIS) in the University of Tokyo Hospital. *J Med Syst* 1996; 20 (4): 197-205.
6. Kuperman GJ, Teich JM, Gandhi TK, Bates DW. Patient safety and computerized medication ordering at Brigham and Women's Hospital. *Jt Comm J Qual Improv* 2001; 27 (10): 509-521.
7. Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. *Arch Intern Med* 2003; 163 (12): 1409-1416.
8. Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, Strom BL. Role of computerized physician order entry systems in facilitating medication errors. *JAMA* 2005; 293 (10): 1197-1203.
9. Horsky J, Kuperman GJ, Patel VL. Comprehensive analysis of a medication dosing error related to CPOE. *J Am Med Inform Assoc* 2005; 12 (4): 377-382.
10. Sowan AK, Mohamed G, Soeken K, Mills ME, Johantgen M, Vaidya V. A comparison of medication administration errors using CPOE orders vs. handwritten orders for pediatric continuous drug infusions. *AMIA Annu Symp Proc* 2006. p 1105.
11. Eslami S, de Keizer NF, Abu-Hanna A. The impact of computerized physician medication order entry in hospitalized patients – a systematic review. *Int J Med Inform* 2008; 77 (6): 365-376.

12. Gorman PN, Lavelle MB, Ash JS. Order Creation and Communication in Healthcare. *Methods Inf Med* 2003; 42 (4): 376-385.
13. Aarts J, Berg M. Same Systems, Different Outcomes – Comparing the Implementation of Computerized Physician Order Entry in Two Dutch Hospitals. *Methods Inf Med* 2006; 45 (1): 53-61.
14. Japan Association of Healthcare Information Systems Industry, JAHIS NEWS vol.14, 2006. <http://www.jahis.jp/en/activity/news.html> (in Japanese) (accessed 27 06 2008).
15. Kimura M, Miyamoto M, Akiyama M, Toyoda K. What We Obtained and What Were the Real Barriers of CPOE and EHR – Lessons Learned From CPOE and RHIO in Japan. Workshop 09, AMIA 2006: 32, Washington DC, USA, November 11, 2006.
16. Sittig DF, Guappone K, Campbell EM, Dykstra RH, Ash JS. A survey of U.S.A. acute care hospitals' computer-based provider order entry system infusion levels. *Medinfo* 2007; 12 (Pt 1): 252-256.
17. Kimura M, Kanno T, Tani S, Satomura Y. Standardizations of clinical laboratory examinations in Japan. *Int J Med Inform* 1998; 48 (1-3): 239-246.
18. Young D. CPOE reduces number of steps in medication-use process. *Am J Health Syst Pharm* 2001; 58 (13): 1170, 1173.
19. Kuperman GJ, Marston E, Paterno M, Rogala J, Plaks N, Hanson C, Blumenfeld B, Middleton B, Spurr CD, Kaushal R, Gandhi TK, Bates DW. Creating an enterprise-wide allergy repository at Partners HealthCare System. *AMIA Annu Symp Proc* 2003. pp 376-380.
20. Kimura M, Shimizu K, Tsuchiya F, Tsuchiya I, Koyama T, Kaihara S, Yashiro N, Iio M. Cross-hierarchy representation of medical knowledge. *Symposium on Computer Application in Medical Care* 87. IEEE Computer Society Press 1987; 812: 207-212.
21. Harrington JJ. IEEE (Institute of Electrical and Electronic Engineers) P1157 Medical Data Interchange (MEDIX): application of open systems to health care communications. *Top Health Rec Manage* 1991; 11 (4): 45-58.
22. http://www.hama-med.ac.jp/index_e.html (accessed 27 06 2008).
22. Kimura M, Tani S, Baatar S, Naito Y, Kanno T, Sakusabe T, Aizawa M. Implementation of multi-vendor DICOM standard image transfer in hospital wide ATM network. *Comput Methods Programs Biomed* 1998; 57 (1-2): 85-89.
24. Health Level Seven, Inc. Standards available at: <http://www.hl7.org> (accessed 27 06 008).
25. Kimura M, Ohe K, Yoshihara H, Ando Y, Kawamata F, Tsuchiya F, Furukawa H, Horiguchi S, Sakusabe T, Tani S, Akiyama M. MERIT-9: a patient information exchange guideline using MML, HL7 and DICOM. *Int J Med Inform* 1998; 51 (1): 59-68.
26. Kimura M. Role of standardization: how to achieve it, and how to cooperate, nationally and internationally? *Comput Methods Programs Biomed* 1994; 43 (1-2): 19-20.
27. Hartleb M, Biernert L, Kochel A. Drug-induced liver damage – a three-year study of patients from one gastroenterological department, *Med Sci Monit* 2002; 8 (4): 292-296.
28. Schachter M. Chemical, pharmacokinetic and pharmacodynamic properties of statins: an update, *Fundam Clin Pharmacol* 2005; 19 (1):117-125.
29. Neuvonen PJ, Niemi M, Backman JT. Drug interactions with lipid-lowering drugs: mechanisms and clinical relevance. *Clin Pharmacol Ther* 2006; 80 (6): 565-581.
30. Barnett GO, Souder DE, Bowie JK, Justice NS. MUMPS – a support for medical information systems. *Med Inform* 1976; 1: 183-189.
31. van Ginneken AM, Smeulders AW, Jansen W, Baak JP, Brooymans I. Design of the diagnostic encyclopedia workstation (DEW). *Comput Biol Med* 1990; 20 (4): 281-292.
32. Cache and XML based Data Exchange Standards, http://www.intersystems.com/cache/whitepapers/xml_standards.html, 2007 (accessed 27 06 2008).
33. Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo Shvo A. HL7 Clinical Document Architecture, Release 2. *J Am Med Inform Assoc* 2006; 13 (1): 30-39.
34. Ferranti JM, Musser RC, Kawamoto K, Hammond WE. The clinical document architecture and the continuity of care record: a critical analysis. *J Am Med Inform Assoc* 2006; 13 (3): 245-252.

Correspondence to:

Michio Kimura
Department of Medical Informatics
Hamamatsu University, School of Medicine,
1-20-1 Handa
Hamamatsu, 431-3192
Japan
E-mail: kimura@mi.hama-med.ac.jp