Searching Electronic Health Records for Temporal Patterns in Patient Histories: A Case Study with Microsoft Amalga

Catherine Plaisant, PhD[1], Stanley Lam[1, 2], Ben Shneiderman, PhD[1, 2],
Mark S. Smith, MD[3], David Roseman, MHA[3], Greg Marchand, MD[3]
Michael Gillam, MD[4], Craig Feied, MD[4], Jonathan Handler, MD[4], Hank Rappaport, MD[4]
[1]Human-Computer Interaction Lab, [2]Dept of Computer Science, University of Maryland,
College Park, MD; [3]ER One Institute, Washington Hospital Center, Medstar Health,
Washington, DC; [4]Microsoft Health Solution Group

Abstract

As electronic health records (EHR) become more widespread, they enable clinicians and researchers to pose complex queries that can benefit immediate patient care and deepen understanding of medical treatment and outcomes. However, current query tools make complex temporal queries difficult to pose, and physicians have to rely on computer professionals to specify the queries for them. This paper describes our efforts to develop a novel query tool implemented in a large operational system at the Washington Hospital Center (Microsoft Amalga, formerly known as Azyxxi). We describe our design of the interface to specify temporal patterns and the visual presentation of results, and report on a pilot user study looking for adverse reactions following radiology studies using contrast.

Introduction

As the use of electronic health records (EHRs) spreads, there are growing opportunities for their use in clinical research and patient care. Queries often have a temporal component. For example "Find all patients who were discharged from the emergency room then admitted again within a week". Another example is "Find patients who had a normal serum creatinine lab test less than 2 days before a radiology test with intravenous contrast, followed by an increase in serum creatinine by more than 50% and of more than 1.0 mg/dl within 5 days after the contrast administration". Currently available user interfaces make possible simple queries such as "Find patients who had a radiology test with contrast and a high value of creatinine", leaving users with the burden of shuffling through large numbers of results in search of matching patients.

Specifying temporal queries in SQL is difficult even for computer professionals specializing in such queries. Database researchers have made progress in representing temporal abstractions and executing complex temporal queries[1, 2, 3, 4], but there is very little research that focuses on making it easy for clinicians and medical researchers to both specify the queries and examine results visually.

Temporal searches are used in many situations, from clinical trial recruitment, clinical research, general patient care (e.g. tracking the application of guidelines) or alarm specification. For example, setting an alarm for patients on Heparin with a precipitous drop in platelet counts (heparin-induced thrombocytopenia) requires specificity around the definition of "precipitous". By querying existing EHR databases, physicians designing the alarm can iteratively test the logic of the alarm (e.g. Should an absolute drop or a relative drop be used as a marker? What is the appropriate time range to use?) and validate it with a large amount of data. Clinicians are always concerned about changes from some baseline state. A blood pressure of 90/60 may be normal for a 25-year old female but may represent severe hypotension in a 65-year old male hypertensive patient whose blood pressure during previous visits was 160/100. In these scenarios, changes from the baseline determine whether or not an intervention should be taken.

We believe that interactive query interfaces allowing researchers and clinicians to explore data that have specific temporal patterns in both numerical and categorical data will dramatically increase the benefits of EHR databases. Comprehensive presentation of the results can then help users see patterns and exceptions in the data they retrieved, and refine their query accordingly. This paper describes our efforts to implement such an interface - called PatternFinder, in a large operational system (Microsoft Amalga, formerly known as Azyxxi) at the Washington Hospital Center (Figure 1).

Related Work

Handling time-related concepts is essential in medicine. A survey[3] lists the many applications that use the history of temporal events, and review the latest contributions to time-aware decision support systems. Research into causality, natural language, and argumentation are hot topics in AI and Medicine research community. However, Augusto observed that there are still opportunities for
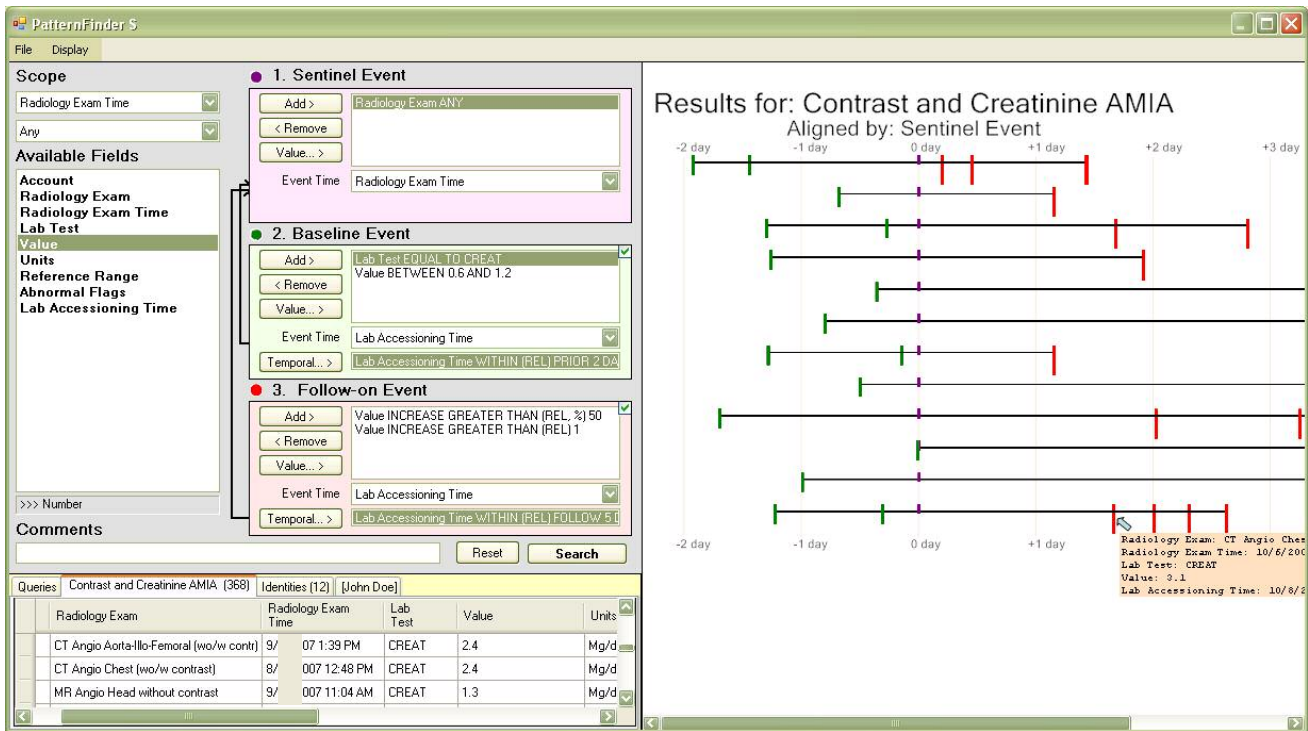
Figure 1 PatternFinder's main screen. The user has specified the following query: "Find patients who had a normal serum creatinine lab test (the baseline event) less than 2 days before a radiology test with intravenous contrast (the sentinel event), followed by an increase in serum creatinine by more than 50% and of more than 1.0 mg/dl relative to the baseline measurement, within 5 days after the contrast administration". The query returned 12 patients. For each patient a timeline shows the events that match the query, aligned by the sentinel event. Events are drawn as color coded tick marks. Baseline events appear to the left of sentinel events, and follow-on events are on the right. Options allow users to display the numerical values as well. Zooming reveals more details. More info at: http://www.cs.umd.edu/hcil/patternFinderInAzyxxi/
NOTE: Medical records numbers and dates have been modified and hidden to produce this figure.

providing much needed search tools. Some research systems provide temporal access languages to support limited visual queries from end-users [5, 6, 7, 8], but many of these suffer the same accessibility difficulties of SQL or they require an understanding of the underlying database structure.

The large body of related computer science work can be grouped into three general areas: time theory, databases, and visualizations. None of the systems discussed in those papers enable query or visualizations of patterns across multiple entities (e.g. temporal cross-patient queries). This paper proposes to combine the power of temporal queries with graphical visualizations, integrated in an operational EHR system, and reports on a pilot study.

Time Theory: Much of the seminal work in computer science relating to time[9,10,11] stems from artificial intelligence, time reasoning, and early natural language processing,.

Databases: Due to the complexity of formulating SQL queries, several approaches have made database query more accessible to a broader spectrum of users, e.g. Query By Example (QBE), the visual query mechanism used in

Microsoft's Access, TSQL2[8], a hybrid between QBE and Extended Entity-Relationship diagrams[12, 13]. MQuery[14] targets various types of streaming data. PROTEMPA allows the interactive specification of temporal abstraction[15].

Visualizations: Chittaro and Combi proposed three alternative visual metaphors for querying temporal intervals[16]. Hibino and Rudensteiner introduced a direct manipulation Temporal Visual Query Language (TVQL)[17] to support Allen's 13 relational primitives. Interestingly, none of the work described above address the visualization of the returned results. However, applications such as TimeSearcher[18], DataJewel[19], KNAVE[6] and LifeLines[20]offer visualizations that cluster results and highlight temporal patterns. LifeLines provides a compact hierarchical timeline visualization for personal histories organized by facets, such as doctor visits, lab tests, and medications[20]. It focuses on a single record and does not offer a query mechanism for discovery across multiple records. Many systems have built on LifeLines, e.g. [21]. Shahar proposes a Knowledge Based Temporal Abstraction model RÉSUMÉ[1].

Again, we emphasize that none of the systems discussed above enable query or visualizations of patterns across multiple entities (e.g. cross-patient query). The implementation of PatternFinder in Amalga builds on early prototypes developed at the Human-Computer interaction Lab[22]. A limited set of temporal filters could be chained together with changes always relative to the previous event. Those prototypes only read small datasets from flat files. We worked with physicians to refine the prototypes and develop a taxonomy of query types. A student project tested the user interfaces and generated recommendations (www.cs.umd.edu/hcil/patternfinder). LifeLines 2 focuses on the exploration of the results once a query has been specified, and our studies demonstrated the benefit of alignment (e.g. on sentinel events)[23].

Description of PatternFinder

Interfacing with Amalga: While Amalga has a wide range of databases, databases are presented to the PatternFinder application using an event-value table data format. An event-value table joins patient information with event information, requiring a field for categorical or numerical event data in addition to a date-time associated with the event. Each row in the event-value table contains information describing at least one event. Although Amalga uses a post-relational data model, the event-value tables used in Amalga are constructed as virtual tables by creating views that join multiple databases and tables together, which can have up to hundreds of fields. Figure 1 shows the interface linked to a data table containing radiology events with contrast and creatinine labs.

Query Specification Interface: To minimize the learning time for Amalga users, the PatternFinder interface has the look and feel of Amalga's existing data filtering tool, to which we added support for temporal functionality, i.e. the ability to specify multiple events, define relationships between events using a time basis, and compare values between events. While we aim to create a very general interface to handle a large number of medically interesting temporal patterns, we chose to focus our initial attention on the relationships of sentinel events, baseline and follow-on events.

Users first define a temporal scope for the query (top left) e.g. choosing a date range for the radiology exam. Next they specify the sentinel event, baseline event and follow-on event of interest in the three separate filter widgets. In the example of Figure 1, the sentinel event is the radiology exam during which intravenous contrast was administered, the baseline event is a lab test result of a normal serum creatinine level, and the follow-on event is another serum creatinine lab test which shows an increase in serum creatinine level of at least 50% relative to the baseline value and of at least 1 mg/dl. The events are linked together using the temporal definition fields and dialog boxes (Figure 2). A date-time field asks users to

select the appropriate database field to use if multiple date-time fields exist in the data source – a likely scenario in Amalga's database tables and views. The addition of these fields in the event definition interface serves as a visual reminder for the steps necessary for fully defining a temporal pattern (Our first iterations of PatternFinder did not include this separate date-time widget, so some users forgot to temporally link the events together). Most operators are relative to other event times or values and we allow users to select the event to be used as reference. Arrows on the left of the filter widgets indicate the chosen temporal connections between widgets (Figure 1).

Figure 2. The temporal relationships use various operators and allow users to specify the reference time (e.g. here the sentinel event or the baseline event).

The 'operators' available in any given event allow users to specify changes in field values between events. For numeric fields, operators include:

- "relative increase greater than X"
- "relative increase greater than X%"
- "relative decrease greater than X"
- "relative decrease greater than X%"
- "less than value in event X"
- "equal to value in event X
- "not equal to value in event X"

For date time fields, new operators include:
- "within X prior to (relative)"
- "within X following (relative)"
- "after X (relative)"
- "before X (relative)"
- "is equal to (relative)"

For text fields, new operators include:
- "equal to value in event X"
- "not equal to value in event X"

Display of Results: PatternFinder for Amalga uses two methods for showing results, a table view that mirrors the existing Amalga "Grid" view and a graphical view which combines aspects from LifeLines[20] and an earlier PatternFinder version using a ball and chain metaphor[22].

On the lower left part of the screen the table view breaks result pairings up and places each event in a single row. User scan review all events, the list of patient IDs that satisfy the query, or view all events for a single selected patient. To increase readability, duplicate rows were eliminated from the table as events could trigger multiple result pairs/triplets. At the right end of each row, additional columns indicate which filter the event was matched to.

On the right side of the screen the graphical display shows all records in the result set, labeled by the patient number (hidden in our figure). For each patient a simplified LifeLines-based timeline summarizes all the events that match the query for that patient. A tick mark is drawn to indicate the time of each event, colored to match the event type, allowing users to see which events relate to each filter of the temporal query. Numerical values can be displayed below the tick marks (in our example the creatinine serum level value would be shown there).

Based on the result of our previous research[23] an "align by" control was added to interactively align the records (either by the sentinel event or the 1st baseline or 1st follow-on event). This action makes the time scale relative to the aligned event. In figure 1, the LifeLines have been aligned by the sentinel event i.e. the time of the radiology exam.

Users can click on a "LifeLine" to expand it and see more details about the matches. Two options are available: 1) a single triplet display to show only one set of events matching the filters (i.e. events that are the closest to the sentinel event) or 2) a display of all the possible triple combinations of events that match the query for that patient (this display is called a ball and chain display[22]. Users can also zoom in to enlarge the display, or zoom out to see more patients and events at once. By hovering over any of the tick-marks from the LifeLines display (or balls in the ball-and-chain display) users can see detailed textual information for that specific event (date and time, test name, units etc.).

Pilot user study

While laboratory-based controlled experiments are effective to measure the performance of interfaces, our expertise in evaluating user interfaces suggests that the complex cognitive skills and domain expert knowledge involved here mean that usability testing and case study methods are more effective in revealing problems and identifying strengths in interfaces such as PatternFinder. Adoption and the value of the findings obtained using the interface eventually becomes the measures of success.

During the past 4 months we worked with 3 senior emergency medicine physicians and a researcher, all from the Washington Hospital Center. In each individual meeting we reviewed the interface, let physicians use it on their own, and discussed needed improvements. The interface was revised between each meeting. We first used a small de-identified dataset provided by the Washington Hospital Center's emergency department, then recently connected to a replicated database mirroring the live database. Being familiar with the Amalga interface and the data available, participants had no problems getting started using the interface and were able to define some temporal queries with limited help. The first iteration of the interface was very generic causing some confusion understanding what was possible. Participants suggested simplifying the interface to tackle simple problems more elegantly. For example, originally the interface had filter labels such as Filter 1, 2 and 3 and no notion of sentinel events. Providing a more guided interface seemed to naturally lead users toward useful queries which were correctly specified. Identifying the needed data table in Amalga remains a challenging step for many users. It may have a significant impact on the overall usability of our interface (even if unrelated to our design) which we hope will be addressed with the development of an interactive data dictionary for Amalga. Participants also identified additional temporal operators that were missing from the initial application: "event count" and "absence of". All physicians were very eager to see the prototype upgraded to the live Amalga system to access the live data, which is now possible. A simple interface allows users to submit multiple queries in batch mode, monitor the progress of long queries, review past queries and results, and attach comments to them.

The case study of adverse reaction to contrast materials illustrates the potential benefit of PatternFinder. A significant decline in renal function after administration of intravenous contrast material is a not uncommon and potentially serious consequence of modern medical care. Its incidence is estimated at 2% - 7% of all administrations of contrast media. Most hospitals are unable to specify in real time (or even retrospectively) which patients and the percentage of patients who have had contrast nephropathy. The ability to identify and track these patients in real time is helpful at several levels. At the individual patient level, the dose of many medications needs to be modified in the face of reduced renal function. Second, at a population level, hospitals are increasingly being asked to provide objective measurement of quality of care. The overall hospital incidence of contrast-induced nephropathy reactions (which typically is not tracked in most hospitals) could become a standardized measure across hospitals if there were a simple way to gather the information. The temporal query methodology we describe here would provide such a way.

Another example of question we investigated is the occurrence of thrombocytopenia, with a query asking for patients with a 50% drop in platelet count within 5 days of heparin administration. While early testing suggests that the query does a good job at identifying populations of interest, PatternFinder needs to link each result directly

to the detail views of the individual patient records (e.g. seeing all the lab results of a patient, instead of just those that match the query). Similarly it will be useful to define an alarm based on a query or to schedule queries to run at regular intervals to track performance metrics.

Conclusion

By bridging the worlds of data bases, user interface design and information visualization, we believe that we can create the next generation of potent visual analytic tools and information technology-enabled work environments. Each participant in our study found ways to use PatternFinder to search for patterns that their current tools could not locate.

References

1. Shahar, Y., Dynamic Temporal interpretation contexts for temporal abstraction, *Annals of Mathematics and Artificial Intelligence*, 1998;22, 159-192.
2. Shahar, Y., Dimension of time in illness: An Objective View, *Annals of Internal Medicine*, 2000; 132 , 1, 45-53.
3. Augusto, J. C., Temporal reasoning for decision support in medicine, *Artificial Intelligence in Medicine*. 2005; 33, 1, 1-24.
4. Stacey, M., McGregor, C., Temporal abstraction in intelligent clinical data analysis: A survey, *Artificial Intelligence in Medicine,* 2007; 39, 1-24
5. Catarci, T., Costabile, M. F. Levialdi, S. and C. Batini, Visual query systems for databases: a survey, *Journal of Visual Languages and Computing*, 1997; 8, 215-260.
6. Cheng, C., Shahar, Y., Puerta, A., and D. Stites, Navigation and visualization of abstractions of time-oriented clinical data, *Stanford University Section on medical informatics technical report* SMI-97-0688, 1997.
7. Cheng C. and Shahar Y., Intelligent visualization and exploration of time-oriented clinical data, *Topics in Health Information Management*, 1999; 20, 15-31.
8. Jensen, C. S. and Snodgrass, R. T., Temporal data management, *IEEE Transactions on Knowledge and Data Engineering*, 1999;11, 36-44.
9. Bruce, B. C., A model for temporal references and its application in a question answer program, *Artificial Intelligence*, 1972; 3, 1-25.
10. Allen, J. F., Maintaining knowledge about temporal intervals, *Communications of the ACM*, 1983; 26, 832-843.
11. Kahn, K. M. and Gorry, A. G., Mechanizing temporal knowledge, *Artificial Intelligence*, 9, 1977, 87-108,
12. Kouramajian, V. and Gertz, M., A graphical query language for temporal databases, in *Proc. of Object-Oriented and Entity Relationship Modeling*: Springer-Verlag, 1995, 388-399.
13. Silva, S. F., Schiel, U., and Catarci, T., Visual query operators for temporal databases, in *International Workshop on Temporal Representation and Reasoning*, 1997, 46-53.
14. Dionisio J. D. N. and Cardenas, A. F., MQuery: A visual query language for multimedia timeline and simulation data, *Journal of Visual Languages and Computing*, 1996; 7, 377-401.
15. Post, A. R., Harrison, J. H.. Protempa: A Method for Specifying and Identifying Temporal Sequences in Retrospective Data for Patient Selection. Am Med Inform Assoc**.** 2007;14, 674-683.
16. Chittaro L. and Combi, C. Visualizing queries on databases of temporal histories: new metaphors and their evaluation, *Data and Knowledge Engineering*, 2003; 44, 239-264.
17. Hibino S. and Rundensteiner, E. A. User interface evaluation of a direct manipulation temporal visual query language, *Proc. of ACM Multimedia*: ACM, 1997, 99-107.
18. Hochheiser, H. and Shneiderman, B., Dynamic query tools for time series data sets: Timebox widgets for interactive exploration, *Information Visualization* 2004; 3, 1, 1-18.
19. Ankerst, M., Jones, D. H., Kao, A., and Wang, C., DataJewel: Tightly integrating visualization with temporal data mining, *ICDM Workshop on Visual Data Mining*, 2003.
20. Plaisant, C., Mushlin, R., Snyder, A., Li, J., Heller, D., and Shneiderman, B., LifeLines: Using visualization to enhance navigation and analysis of patient records, *American Medical Informatics Association 1998 Annual Fall Symposium*, Bethesda MD, 1998, 76-80.
21. Bade, R., Schlechtweg, S., and S. Miksch, Connecting time-oriented data and information to a coherent interactive visualization, *Proc. ACM Conf. on Human Factors in Computing Systems (CHI 2004)*, ACM, 2004, 105-112.
22. Fails, J., Karlson, A., Shahamat, L., and Shneiderman, B., A visual interface for multivariate temporal data: Finding patterns of events over time, *Proc. IEEE Symposium on Visual Analytics Science and Technology*, IEEE, 2006, 167-174.
23. Wang, T., Plaisant, C., Quinn, A., Stanchak, R., Shneiderman, B., and Murphy, S., Aligning temporal data by sentinel events: Discovering patterns in electronic health records, *Proc. ACM Conf. on Human Factors in Computing Systems (CHI2008)*, ACM, 2008, 457-466.