KNAVE-II: A Distributed Architecture for Interactive Visualization and Intelligent Exploration of Time-Oriented Clinical Data

Yuval Shahar, Dina Goren-Bar, Maya Galperin, David Boaz, and Gil Tahan

Department of Information Systems Engineering,
Ben Gurion University, Beer Sheva 84105, Israel
{yshahar,dinag,gmaya,dboaz,gilta}@bgumail.bgu.ac.il

ABSTRACT

Interpretation and exploration of longitudinal clinical data is a major part of diagnosis, therapy, quality assessment, and clinical research, particularly for chronic patients. KNAVE-II is an intelligent interface to a distributed architecture specific to the tasks of query, knowledge-based interpretation, summarization, visualization, interactive exploration of large numbers of distributed time-oriented clinical data and dynamic sensitivity analysis of these data. The web-based architecture enables users (e.g., physicians) to query, visualize and explore clinical time-oriented databases. Both, the generation of context-sensitive interpretations (abstractions) of the time-stamped data, as well as the dynamic visual exploration of the raw data and the multiple levels of concepts abstracted from these data, are supported by runtime access to domain-specific knowledge bases, maintained by domain experts.

KNAVE-II was designed according to a set of well-defined desiderata. The architecture enables exploration along both absolute (calendar-based) and relative (clinically meaningful) time-lines. The underlying architecture uses standardized vocabularies (such as a controlled dictionary for laboratory tests and physical observations), and predefined mappings to local data sources, for communication among its various components. Thus, the new framework enables users to access and explore multiple remote heterogeneous databases, without explicitly knowing their local structure and vocabulary, through a filter of a set of task-specific knowledge bases. The complete architecture has been implemented and is currently evaluated by expert clinicians in several medical domains, such as oncology, involving monitoring of chronic patients.

Background

Management of patients, especially chronic patients, requires collection, interpretation and exploration of large amounts of time-oriented data. The task of creating interval-based concepts (abstractions) from time-tamped raw data is called temporal abstraction [Shahar, 1997]. An example is abstraction of periods of Bone-Marrow Toxicity (as defined in a particular context) from raw individual hematological data (Figure 1).

![Figure 1: Temporal abstraction of post bone-marrow transplantation data. Raw data are plotted over time at the bottom. External events and the abstractions computed from the data are plotted as intervals above the data. — — = an external event (medical intervention); • = platelet counts; Δ = granulocyte counts; — — = a context interval; — = an abstraction (derived concept) interval; BMT = bone-marrow transplantation (an event); PAZ = a therapy protocol for treating chronic graft-versus-host disease (CGVHD), a complication of BMT; B[n] = bone-marrow–toxicity grade n, an abstraction of platelet and granulocyte counts.](image-url)
The ability to automatically create interval-based abstractions of time-stamped data, in particular in the medical domain, has multiple implications. **Data summaries** of time-oriented electronic data have an immediate value to a human user, such as to a physician scanning a long patient record for meaningful trends [Downs et al., 1986; De Zegher-Geets, 1988]. **Temporal abstractions** support recommendation and explanation of recommended actions by intelligent decision-support systems, as well as monitoring of plans (e.g., therapy) during execution [Musen et al., 1996], and are a useful representation for intentions of clinical guidelines [Shahar et al., 1998]. Meaningful time-oriented contexts enable generation of context-specific abstractions, maintenance of several interpretations of the same data within different contexts, and certain hindsight and foresight [Shahar, 1998]. Domain-specific, meaningful, interval-based characterizations of time-oriented data are a prerequisite for effective visualization and exploration of these data [Cousins and Kahn, 1991; Shahar and Cheng, 1999, 2000].

A problem solver implementing a formal methodology for knowledge-based temporal abstraction [Shahar, 1997], called **RESUME**, has been implemented and evaluated in several clinical domains [Shahar and Musen, 1996]. The knowledge required by the system is acquired from medical domain experts [Shahar et al., 1999]. A temporal mediation architecture [Nguyen et al., 1999]; has been developed to integrate the data bases, knowledge bases, and the computational temporal-abstraction module.

One of the ways to support care providers performing tasks such as diagnosis, therapy, and research is by supplying them with the technology for on-the-fly visualization, interpretation and exploration of the data and of the higher level, knowledge-based, concepts that can be derived from these data.

Visualization and exploration of information in general, and of large amounts of time-oriented data in particular, is essential for effective decision making. Larkin and Simon [1987] have demonstrated that the usefulness of visual representation is mainly due to (1) the reduction of logical computation through the use of direct perceptual inference, and (2) the reduction of necessary search for information through the use of efficient graphical representation.

Researchers in the areas of visualization of clinical time-oriented data [Cousins and Kahn, 1991; Powsner and Tufte, 1994], have developed useful visualization techniques for static presentation of raw time-oriented data and for browsing information. An excellent treatise on visualization is the series of books by Edward Tufte on methods to display information [Tufte 1990, 1997]. The Lifelines project [Piaisant et al., 1996, 1998] developed an intuitive visualization of historical events and data, demonstrated also in a medical domain, although it does not include any domain-specific abstraction knowledge and no abstraction mechanisms.

In recent years, researches have investigated various techniques for information visualization, including conceptual maps, radar maps, tree maps (conetree/camtree maps, hyperbolic tree), Kohonen maps, fish eye views and dynamic queries interfaces and more [Card et al., 1999; Chi, 2002; Hearst, 2000; Shneiderman, 1994; Shneiderman et al., 2000], as well as using well-known statistical and graphical methodologies, such as three-dimensional representations, scattergrams, pie charts, bar charts, and their derivative techniques. These display methods, however, typically do not focus on visualization of domain-specific temporal abstractions and on the issue of interactive manipulation and exploration of the data and multiple levels of its abstractions, using domain-specific knowledge. Typically, these capabilities have been omitted, because they require a formal, domain-independent representation of the domain-specific temporal-abstraction knowledge, considerable effort in modeling the visualized domain, and availability of computational mechanisms for creation of the abstractions.

**Desiderata for Interactive Exploration of Time-Oriented Clinical Data**

The desiderata for provision of a service for interactive exploration of time oriented clinical data include the following requirements:

D-1. **Accessibility**—A modular, scalable architecture to enable the application of diverse (and preferable distributed) types of knowledge to the same data and different data bases to the same knowledge; the architecture should support access to both the data and the abstractions derivable from it.

D-2. **Visualization of raw data and abstractions**—Effective visualization and exploration, which should include both raw data and the abstract concepts derived from those data.

D-3. **Temporal granularity**—The visualization should support interactive exploration of time oriented data at different levels of temporal granularities.

D-4. **Absolute and Relative time lines**—The system should support both a calendar-based time line as well as a relative time line, which refers only to clinically significant events (e.g. start of therapy).

D-5. **Intelligent exploration of raw data and abstractions**—Effective exploration of both raw data and their abstractions, using meaningful domain-specific semantic relations.
D-6. **Explanation**—The system should enable an explanation of abstractions from data and knowledge.

D-7. **Statistics**—The system should provide statistics on both the raw data and the abstracted concepts.

D-8. **Search and Retrieval**—The system should support easy and fast search and retrieval of clinically significant concepts.

D-9. **Dynamic Sensitivity Analysis**—The system should include capabilities for interactive exploration of the effects of simulated hypothetical modifications of raw data on the derived concepts.

D-10. **Clinical-Task support**—The system should be customizable for a specific clinical task (e.g. monitoring of diabetic patients).

D-11. **Collaboration**—The system should support collaboration between different clinicians and researchers on the same patient data and abstracted knowledge.

D-12. **Documentation**—The system should support documentation of the exploration process.

**The KNAVE-II Architecture**

In previous research, we have implemented a stand-alone prototype module of the visualization service, called knowledge-based navigation of abstractions for visualization and explanation (KNAVE). Preliminary assessments of KNAVE in the oncology domain were highly encouraging, and demonstrated the feasibility of the whole architecture and in particular of the knowledge-based exploration concept [Shahar and Cheng 1999, 2000].

We have recently developed a preliminary advanced prototype module, KNAVE-II, a significantly enhanced intelligent interface designed to fulfill all the above mentioned desiderata. In this section, we explain in detail how the KNAVE-II architecture supports all desiderata (D-1 to D-12).

In order to support the **accessibility** requirement (D-1), a new knowledge-based distributed temporal-abstraction mediation architecture, IDAN [Boaz and Shahar, 2003a], was designed and implemented. IDAN uses a modern version of the RESUME problem solver, ALMA [Boaz and Shahar, 2003b]. The modular architecture includes automated acquisition of domain-specific temporal-abstraction knowledge, a computational temporal-abstraction mechanism using that knowledge, a data-access service that accesses time-oriented databases, and controlled-vocabulary servers. The modular architecture includes multiple knowledge bases and time-oriented databases (Figure 2).

A full scalable distributed architecture requires the capability of **remote connectivity** to diverse data bases, knowledge bases, vocabularies and algorithms to enable the application of types of knowledge to the same data and different data bases to the same knowledge. A **configuration** service screen enables users of a KNAVE-II client to select, at the beginning of an exploration session, the desired data base, knowledge base, and temporal-abstraction service.

![Figure 2: The knowledge based distributed architecture.](image-url)
To solve the visualization desiderata, we use an intelligent user interface. Intelligent user interfaces are knowledge-based interfaces that mediate between person and machine to increase the ease and effectiveness of user interactions [Goren-Bar, 1999, 2001]. KNAVE-II is a knowledge-based, interactive visualization and exploration intelligent user interface. The interface is used to explore a single patient record or a set of such records. Figure 3 shows the main interface of a KNAVE-II visualization and exploration client. KNAVE-II enables interactive, dynamic exploration by the user of raw data and their abstractions (D-2). Figure 3 demonstrates our current design for the integration of a knowledge browser (which reflects the contents of the domain’s temporal-abstraction knowledge base, that is, its temporal-abstraction ontology) and data panels (which show the contents of the database or of the results of a temporal-abstraction process applied on such contents).

Figure 3: A view of an individual patient’s data in the KNAVE-II prototype (in, an oncology domain). On the left hand side, a browser to the clinical domain’s ontology, coming from the ontology knowledge base, is shown. The user selects a raw data (3rd, 5th and 6th panels from the top) or abstract concept (1st, 2nd and 4th) from the ontology tree or by asking a query from the left hand bottom panel, which is then retrieved or computed on the fly and displayed as a panel on the right hand side. Operators represented as icons, in each panel, enable the user to perform actions such as: (a) The time-synchronization function [pin-shaped] icon, enables to synchronize the display of the panels according to the specified time in the selected panel (b) query the knowledge used to derive the concept through the “kb” icon, (c) add statistics clicking on the stats icon about raw and derived concepts (see the statistics displayed on 3rd and 1st panels from the top, respectively) (d) semantically explore the concept clicking on the cross icon (e) and the right and left direction arrows enable to skip to the nearest period (at the selected direction) data was found. The set of widgets that controls all panels: (f) the random granularity zoom, which enables slide-bar manipulation of any desired temporal granularity; (g) the calendaric-range zoom, which enables zooming into a time range, by selecting the start and end time points using a calendar; (h) the patient-selection box, which selects the current patient-record to explore; (i) the statistics button, which controls whether to add or to remove statistical information; (j) the absolute/relative time line, which enables to set a specific event (such as a medical intervention) as a date of reference to all the other displays. (k) The search and retrieval service enables lexicographic search. The user types a string in the entry window. KNAVE-II retrieves all the related concepts. Clicking on the Find Similar checkbox triggers KNAVE-II to retrieve also similar concepts without requiring the user to have prior knowledge about the exact form a concept appears in the ontology knowledge base. The concepts can be ordered according to their type and the related context, and then opened for further visualization.
KNAVE-II also supports the exploration of time oriented data at different levels of temporal granularities (D-3). KNAVE-II implements five operators (zoom-in functions) for manipulating temporal granularity: (a) a random granularity zoom, which enables specification of any desired temporal granularity; (b) a calendaric-range zoom, operates a calendar function by specifying the start and end time points to zoom-in into a specific absolute time range; (c) a single panel zoom, opens a particular concept or data panel in a separated sub-window; (d) a time-sensitive zoom, to select a specific predefined period of time within the timeline of a particular panel e.g., “August 1995” and (e) a content-based zoom to mark-up specific contents in the panel and then zoom into the temporal range implicitly determined by these contents, whether within a complete temporal-granularity unit or not. Figure 4 demonstrates the time-sensitive zoom and the content-based zoom.

Changing dynamically the point of view from an absolute (calendar-based) time line to a relative time line (D-4) is another KNAVE-II innovative capability. The relative time line is set by identifying clinically significant events (e.g., start of therapy, birth of the child) which serve as a date of reference to all the other displays. Once the relative time-line was selected the time display will change to +/- time units starting from that event, based on the time granularity selected (hours, days, months, years). The user can select the event to be used as the zero-time reference, through access to a predefined list of events, defined within the knowledge-base, that can be used as reference points. (Figure 5).

Figure 4- The time-sensitive zoom enables users to zoom into a specific predefined period of time, by clicking on a specific temporal granularity button within the timeline of a particular panel, e.g., “August 1995”. The content-based zoom enables users to mark specific contents in the panel whether within a complete temporal-granularity unit or not (see shadowed area), and then zoom into the temporal range implicitly determined by these contents.

Figure 5- Absolute/relative time lines in KNAVE-II. Once the relative time-line was selected the time display will change to +/- units starting from that event. The selection of the time reference event can be done by direct manipulation or by selecting a predefined reference event from the knowledge-base, in which case KNAVE-II will show the nearest event enabling direct browsing between events (in the case that there were more than one such points in the patient’s record).
Exploration of raw data and abstract concepts (D-5) includes navigation along semantic links in the domain’s temporal-abstraction ontology, such as abstracted-from relations, using the semantic explorer (Figure 6a). During exploration, the user is able to obtain context-sensitive explanations (D-6) to questions such as “From which data is this concept abstracted?” and “What classification function defines this abstraction?” (Figure 6b). For any domain, the semantics of the query, visualization, and exploration processes are the same, since these processes use the terms of the domain-independent knowledge-based temporal-abstraction ontology. However, the exploration operators use, for any domain, the domain-specific temporal-abstraction properties of that particular domain. The result is a uniform-behavior, but context-sensitive (with respect to the knowledge) visualization and exploration interface in all time-oriented domains.

To support clinical research, it is imperative to provide several types of descriptive statistics as part of the interactive visualization and exploration. Statistics (D-7) in KNAVE-II can be computed and displayed for either raw data or abstracted parameters. The computation of statistics is, sensitive to the particular time window displayed in each panel, and thus changes dynamically when the contents of the panel are changed. Default statistics for raw data types include descriptive statistics such as mean, maximum, minimum, standard deviations, etc. (see 3rd panel from the top in Figure 3). In the case of abstract data types, the default statistics displayed are a detailed distribution of the duration of each value of the abstraction (e.g., GRADE-II bone-marrow toxicity) (see 1st panel from the top in Figure 3).

KNAVE-II supports easy and fast search and retrieval (D-8) of ontology-based clinically-significant concepts. The search supports easy ordering of concepts according to the type and the related context, opening the found concept in the knowledge browser, and (optional) opening the matching panel for further visualization and exploration (see Figure 3).

The exploration supports, among other features, dynamic simulation of hypothetical raw data or domain-specific knowledge. The user is able to simulate the effect of modifying the data or the knowledge, thus adding a dynamic sensitivity analysis capability (D-9) (“What-if dynamic simulation) by changing or adding values in a specific panel. The modified values are kept in the cache and do not affect the real patients’ data in the data base. The display reflects the computational implications of these modifications. Exploiting the direct access to the domain’s temporal-abstraction ontology and to the temporal abstraction server will provide these explanations (Figure 7).

**Figure 6:** Exploration of data and knowledge in the KNAVE-II prototype. (a) The semantic explorer, evoked by clicking on the exploration (cross icon) button of the panel (see Figure 3). The user uses the semantic relationships of a concept, which depend on its type (e.g., abstracted-from, abstracted-into, sibling-argument, generated-context, and is-a, in the case of a raw or abstract data type) to navigate to other concepts semantically related to the original concept. (b) knowledge-based explanation, evoked by clicking on the knowledge button (kb) in a panel (see Figure 3). The user queries the temporal-abstraction knowledge that was used to derive a specific displayed concept.
Figure 7: Dynamic sensitivity analysis enables the following actions: adding the hypothetical raw data, modifying the existing data by changing their values and time-stamps; or deleting selected data. The apply button triggers the computation of abstractions derived from the modified raw data, while the undo button cancels the modifications and their effects.

Clinical-Task support (D-10) is achieved by enabling the physicians and medical researchers to select several raw-data and abstract-concept panels that are all related to a specific clinical task (e.g. monitoring of diabetic patients). KNAVE-II enables customization of the displays for a specific clinical task by opening the selected panels and saving them as a profile (e.g., a diabetes profile) in the knowledge base through the profile-save icon in the top-level menu. Clicking on the profile-load icon, another icon in the top-level menu, retrieves all the profiles, enables selection of a specific profile, and applies it to the current patient record.

Clinicians and researchers usually like to consult or share the result of the exploration of data and abstractions with colleagues. KNAVE-II enables collaboration (D-11) by saving the selected data and abstractions of a particular exploration to a special exploration file format, through another one of the top-level menu icons. An exploration file can be saved to a shared directory or sent by email to a collaborating colleague. The collaborator can open the file, and explore the same data, starting from the same point in which the image of the exploration was saved. She also can add other raw-data and abstract-concept panels that seem to be relevant to the case in discussion, and continue the visualization and exploration session (and even send it back).

Standard clinical, research and administrative procedures require documentation (D-12) of the patient's clinical data to the patient's file, thus showing the exploration that supported a clinical decision. Clicking on the camera icon at the top-level menu enables taking a snapshot of the exploration process. Another icon enables printing the exploration for further reference.

Discussion and Future Work

We have introduced KNAVE-II, an advanced, fully implemented prototype of a visualization interface which is able to use the advanced features of an innovative distributed architecture for intelligent visualization and exploration of time-oriented data. The computational architecture supports the generation and exploration of context-sensitive interpretations (abstractions) of the time-stamped data in terms of domain-specific concepts and temporal patterns.

We have started by listing multiple desiderata for supporting the needs of care providers who need to browse large amounts of time-oriented clinical data and reduce the cognitive burden involved in that task. Our key insight is the addition of an interactive intelligent interface supported by an architecture that integrates knowledge and data. We think that one of the needs of a framework for intelligent data analysis is the capacity to explore dynamically both raw data and the interpretations derivable from these data (using domain-specific knowledge); and perform this task in a scalable fashion.

Initial applications of KNAVE-II and its supporting knowledge based temporal-mediation architecture, to a large data base of patients monitored several years after bone marrow transplantation (BMT), have resulted in highly encouraging preliminary impressions. A formal evaluation is under way.

Future work will involve:

A. Visualization and exploration of large groups of patients selected according to dynamic criteria. For instance, exploration of the duration-distribution of bone-marrow toxicities of all patients who received a certain therapy protocol.

B. Additional capabilities for adding customized plug-in visualization modules that extend the capabilities of KNAVE-II.

C. Enhanced capabilities for search and retrieval that use standardized vocabularies (e.g by connecting to the UMLS meta-thesaurus).

D. Supporting on the fly specification (at exploration time) of new temporal patterns, possibly saving them as part of the user’s profile.

Acknowledgements

This research was supported in part by NIH award No. LM-06806. We thank Drs. Mary Goldstein, Susana Martins, Lawrence Basso, Herbert Kaizer, Aneel
Advani, and Eitan Lunenfeld, for their useful coments regarding the KNAVE-II interface.

Bibliography


De Zegher-Geets, I.M. IDEFIX: Intelligent Summarization of a Time-Oriented Medical Database. M.S. Dissertation, Program in Medical Information Sciences, Stanford University School of Medicine, 1987; also Knowledge Systems Laboratory Technical Report KSL–88-34, Department of Computer Science (Stanford University, Stanford, CA, 1988).


