Visual exploration and cohort identification of acute patient histories aggregated from heterogeneous sources

Rune Sætre*, Øystein Nytrø*, Stein Jakob Nordbø†, Aslak Steinsbekk‡
* Department of Computer and Information Science, Sem Sælands vei 9, NO-7491 Trondheim, Norway
Email: {satre,nytroe}@idi.ntnu.no
† Software Developer, Aaengveien 12D, NO-7020 Trondheim, Norway
Email: sj@steinjakob.net
‡ Department of Public Health and General Practice, Post box 8905 MTFS, NO-7491 Trondheim, Norway
Email: aslak.steinsbekk@ntnu.no
Norwegian University of Science and Technology (NTNU)

***Abstract*—How can we use information visualization to support retrospective, exploratory analysis of collections of histories for patients admitted to acute care? This paper describes a novel design for visual cohort identification and exploration. We have developed a tool that integrates multiple, heterogeneous clinical data sources and allows alignment, querying and abstraction in a common workbench.**

This paper presents results from two projects and a review of related work in the field of information visualization including both presentation and interactive navigation of the information. We have developed an interactive prototype and present the visualization aspect of this prototype and a brief demonstration of its use in a research project with a large cohort of patients.

The prototype represents and reasons with patient events in different OWL-formalizations according to the perspective and use: One for integration and alignment of patient records and observations; Another for visual presentation of individual or cohort trajectories.

Health researchers have successfully analyzed large cohorts (over 100,000 individuals) using the tool. We have also used the tool to produce interactive personal health time-lines (for more than 10,000 individuals) on the web. Utility, usability and effect have been tested extensively and the results so far are promising.

We envision that clinicians who want to learn more about groups of patients and their treatment processes will find the tool valuable. In addition, we believe that the visualization can be useful to researchers looking at data to be statistically evaluated, in order to discover new hypotheses or get ideas for the best analysis strategies. Our main conclusion is that the tool is usable, but it can be challenging to use for large data sets.

I. INTRODUCTION

Norway had an early adaption of electronic health records, which means that there are databases containing long and comprehensive patient histories. There are several complementary ways that knowledge can be extracted from these databases. 1) By reading the narrative record directly. 2) By statistical indicator analysis. 3) By Knowledge-Discovery and Data Mining (KDD). In this project, we are interested in a fourth way: By visualizing information from the database in a way that enables clinicians to use his or her own visual processing system to get an overview and discover features in a collection of patient histories.

A. Information visualization

Most current tools for inspection of treatment histories focus on presenting information about the individual patient. We believe that interesting knowledge can be discovered by investigating *groups* of patients, and that information visualization is well suited for this task. Hence, our research question is as follows:

> *How can we use information visualization to support retrospective, explorative analysis of collections of patient histories?*

Information visualization is the use of interactive visual representations of abstract data to amplify cognition [1, Foreword]. This amplification serves as an extension of our limited short-term memory, allowing a complex task to approached by analysing visual patterns. Also, a visualization may reveal artifacts in the data: In one instance, statisticians were analysing a data set for a long time[1] before one of them realised that they had mixed up part of the data – a fact that was discovered using a visualization [2]. From this example, it seems that visualizations have the ability to generate insight.

On machines, interaction is also an important part of visualization. Extending the static diagram with interactive operations allows the analyst to see different views of the same data, helping to form a more complete understanding. Comparisons can be composed, bringing different parts of the chart together, and the level of detail can be varied dynamically. This allows the analyst to build up a mental model of the data, using the visualization as an extension of his short-time memory.

B. Visualization of collections of patient histories

We believe that the application of information visualization techniques to collections of patient histories may be a valuable

[1] According to [2], the data set was included in "a number of statistics texts".

addition to the current repository of electronic health record (EHR) tools: Current EHR tools constitute a computerization of the traditional health record, allowing the user to view and manipulate the individual record on the screen, mainly in text-form. Projects such as the *LifeLines* project [3], [4], [5] aim to enhance usability of the EHR through information visualization. The emphasis is on the individual patient and it can give valuable insights into each unique case.

However, more can be learnt from visualizing collections of cases. This paper addresses the issue of designing an information visualization with accompanying explorative tools, facilitating retrospective investigation of a collection of patient histories. To do so, some questions have to be explored:

- How can suitable information visualization techniques be applied? This includes both user interface design and usability considerations.
- What are the interesting properties of patient histories, and how can meaningful groups of these be extracted? To answer this question, relevant operations and efficient algorithms must be proposed and analyzed. Also, database-technical issues must be considered.
- How can the visual representations of the different elements in the histories be designed to harness the capabilities of the visual processing system of the user? This question is related to cognitive psychology.

Although each of these questions can be approached separately, a workable prototype must be able to present an integrated solution to these questions. To do so we base this paper on two previous projects [6], [7] and a review of the literature on visualization. This is presented in the background. We then present our solution and give an example of its use in a research project with several thousand real patients.

Figure 1 shows the main window of the final prototype, along with a brief explanation of the elements in it.

II. Background

This section introduces our former work and describes the theory and research related to the interactive visualization design proposed in this paper. Finally, other related visualization designs and tools are presented.

A. Previous Projects

1) Visualization of diagnosis histories: The project NSEPter is a visualization prototype using directed graphs to portray collections of patient histories. The only information from the EHR that was utilized, was the diagnosis codes for each patient. NSEPter had the following functionality [6]:

- Each history was laid out on a horizontal line, and each diagnosis code was represented by a node, with an edge between nodes representing diagnoses adjacent to each other in the history.
- The system was capable of searching for diagnosis instances based on a regular expression over International Classification of Primary Care (ICPC-2) codes[2]. This search could be used to hide or show individual nodes, or it could operate on the level of histories, based on the presence or absence of a given code.
- Nodes could be *merged*: The users specified a regular expression over the ICPC codes, and the application merged nodes with codes matching the given expression into one. This was performed serially from the beginning of the histories, so that the first occurrence of a node from one history was merged with the first from all the other histories, the second was merged with the second, and so on. From each merged node, the process could be recursively applied to neighbouring nodes in both directions, in a hope that the histories would exhibit similar patterns before or after an important event. Common edges between merged nodes were scaled according to the number of histories exhibiting the transition in question.
- NSEPter had a plug-in architecture in which filters and visualization engines could be interchanged, all operating on the same data model.

In Figure 2(a), NSEPter is showing a graph of diabetes patients. Here, the thicker lines indicate that several patients follow the same path before and after the diabetes code, T90, which in this case is the first occurrence in all the histories.

This NSEPter prototype had several weaknesses: First, in the transformation from histories to graph representation, the dimension of time was lost. This made medically vital information absent in the visualization, and it was no way of deciding if the time that passed between two events was half an hour or a year.

Second, the graphs quickly became crowded and virtually unreadable, and they used a lot of screen space. The same can be seen in several other systems and in other visualization domains as well (e.g. Figure 1 in [8]). If one zoomed out to see the big picture, the visualization was basically a web of edges, and with larger zoom factors, context was lost, and it was difficult to determine what one was looking at (Figure 2(b) illustrates this). Third, the merging algorithm was not very noise-resilient. It would miss an opportunity to merge nodes if two histories differed in one single position. Moreover, the order in which the histories were merged, mattered.

2) Directed graphs and merging similar paths: The second project focused on improving the basic idea of using directed graphs and merging similar paths [7]. This project employed alignment methods and different measures to reduce the amount of noise. In addition, we calculated abstractions over sequences of diagnosis instances and mined for relations between the diagnosis codes themselves. From this project, we gained experience in processing the available data. In particular, the filtering and selection techniques presented in this paper build on this experience. Among the future work listed in the conclusion of [7] were goals such as 1) incorporate more information in the visualization and 2) better exploit the dimension of time.

B. Literature on information visualization

The human brain possesses a considerable capacity of pattern recognition through visual processing, while the ability

[2]http://www.who.int/classifications/icd/adaptations/icpc2/en/

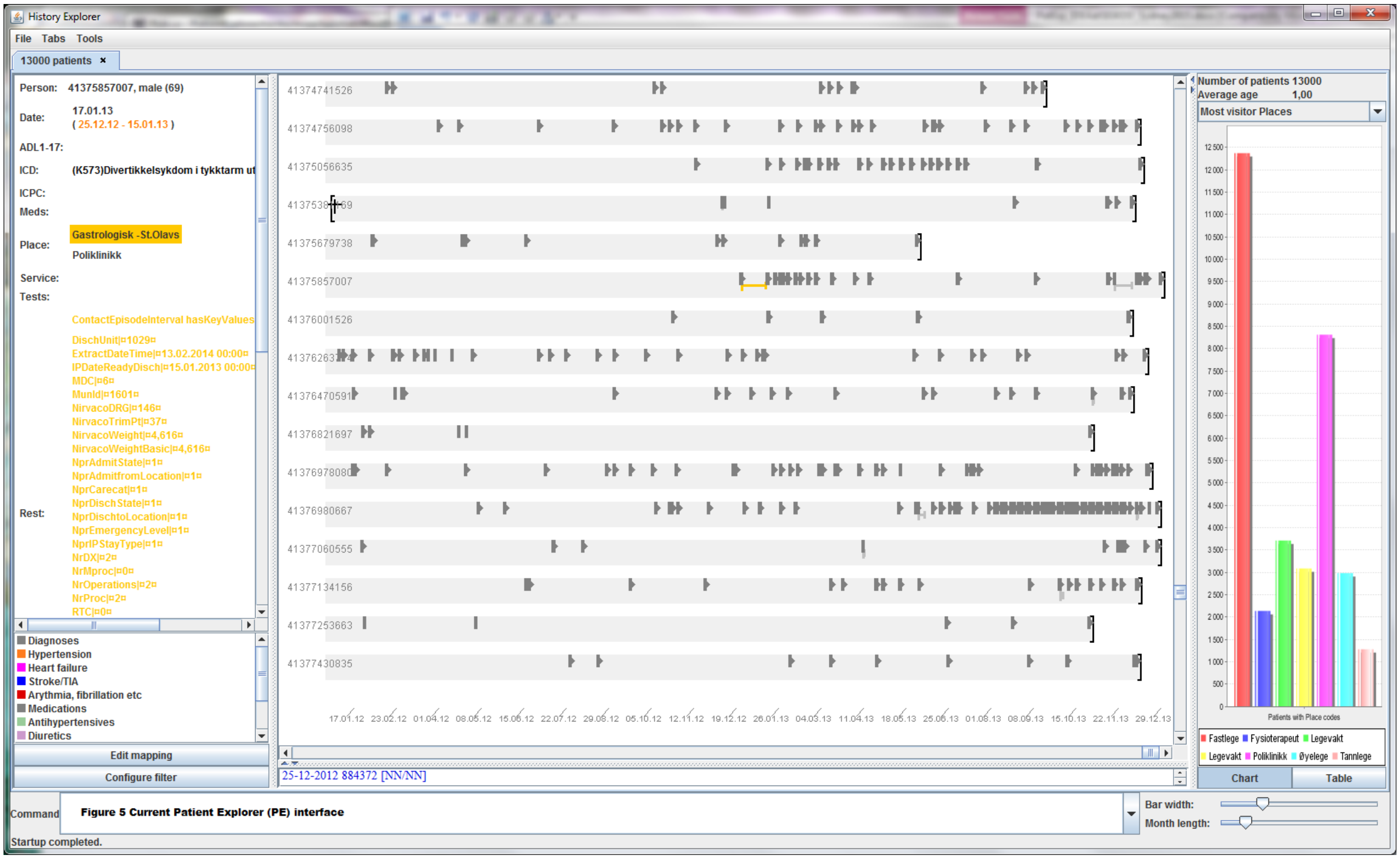


Fig. 1. The visualization prototype. Each gray bar in this figure constitutes a patient history, with small rectangles and arrows indicating diagnoses and blood pressure measurements, respectively. The colors in the visualization show different classes of medication. On the left-hand side and bottom of the window, there are dynamic displays showing detailed information about the history content under the mouse cursor.

for processing large amounts of text or numbers is limited in comparison. For computers, the situation is opposite: While possessing the ability for processing huge amounts of data, the algorithms for pattern-recognition are limited, confined to finding one or a few types of solutions, and often they are sensitive to noise and errors, as mentioned (and ignored) in [9]. Other benefits of human cognitive abilities are access to creativity and the availability of domain knowledge, conscious or not, stored in the heads of users [10].

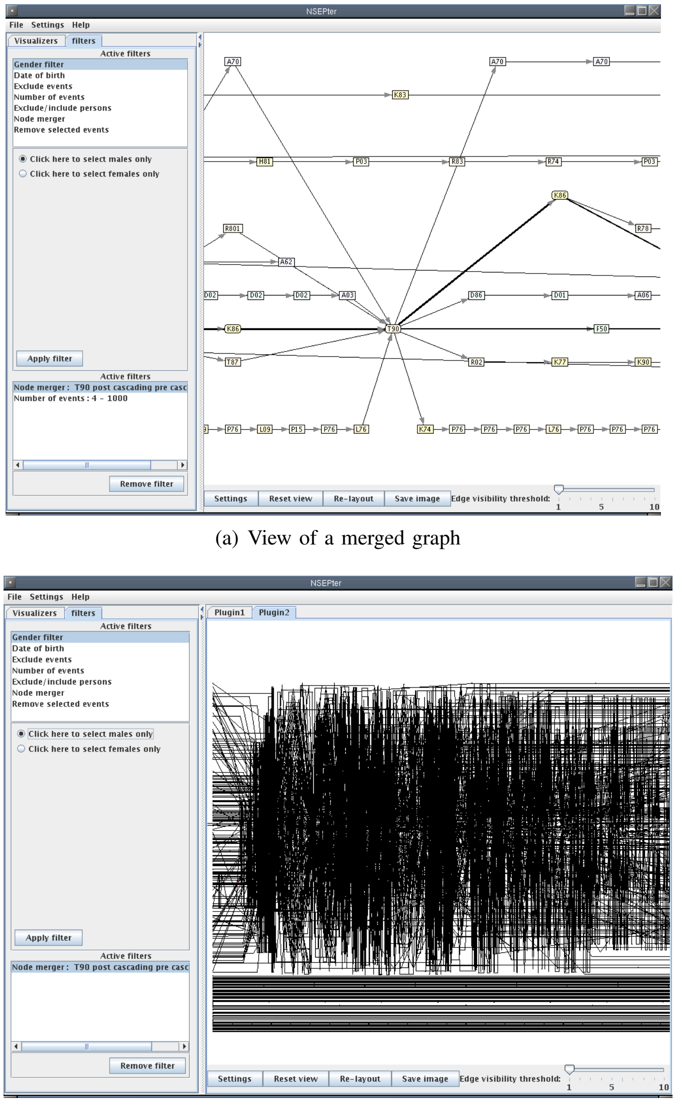
Using the computer for number crunching and construction of a visual representation, the strengths of man and machine are combined for extraction of interesting properties of the data being investigated. The high availability of computing power and high-resolution displays makes information visualization attractive and feasible, even on a low-end home computer.

When designing an information visualization, knowledge of how different visual structures are perceived serves as a useful guideline in making the visualization easy to learn and understand. This includes choosing graphical primitives that can be quickly identified, and arranging them for the best possible exploitation of the perceptual system. If the visualization is well crafted, searching for specific information becomes easy since it can be done preattentively, see 1) below, and obvious patterns are quickly revealed. In addition, cognitive limitations must be observed: When an image becomes too complex, it becomes difficult to read and interpret. Also, the short-term memory for graphical elements is quite limited. In fact, experiments indicate that we are not able to detect even large differences when a display changes abruptly [11].

1) Preattentive processing: Sometimes, features seem to “pop out” of an image. For example, when trying to locate the red circle among the blue ones in Figure 3, the differing color is picked out by the visual processing system before the signal reaches the centre of attention – it is performed *preattentively*. The time used to process the visualization (search for the red circle) is independent of the number of distracting elements (blue circles). There are a number of features that can be processed preattentively; for example is the same effect evident when searching for circles in a figure with many squares or other angled forms [12].

On the other hand: Searching for a red circle in a figure with many blue circles and red squares cannot be performed preattentively, and the time to do so increases linearly with the number of distracting elements ([12], [1]). This is called a *conjunction search*, since more than one property need to be processed in order to identify the target: It has to be red, and it has to be circular. In general, conjunction search is not preattentive, but there are important exceptions [1].

Since preattentive processing is much faster than its counterpart, choosing a suitable visual encoding is important for the efficiency of the resulting presentation. This includes choosing good colors and distinct forms, and avoiding the need for



(a) View of a merged graph

(b) Zoomed-out view of a merged graph

Fig. 2. The NSEPter prototype showing (a) a small graph, merged around the first incidence of diabetes, (b) several hundred patients, showing the entire graph.

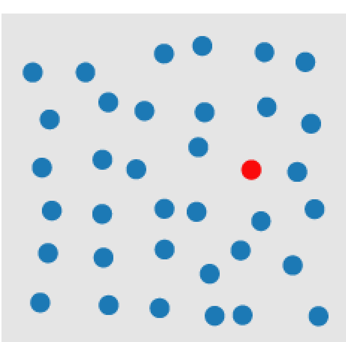


Fig. 3. Find the red circle [12]

conjunction search.

2) Choice of shapes: Preattentive concerns should also be addressed when selecting the shapes. Shapes that are preattentively processed are often simple shapes, and the different shapes that are chosen should be sufficiently distinct. Ware lists different features that are preattentively processed (from [1]): Line orientation, Line length, Line width, Line co-linearity, Size, Curvature, Spatial grouping, Blur, Added marks, Numerosity, Color hue, Color intensity, Flicker, Direction of motion, 2D position, Stereoscopic depth and Convex/Concave shape from shading.

Other related recent health-data visualization approaches are reviewed in [13], [14] and [15], with a concrete example in [16].

C. Literature on interactive computer graphics

Information visualization techniques do not only concern the construction of a static diagram, but also how to interact with the visualization. Using these techniques, parameters can be varied interactively to achieve a view of the data highlighting the properties the analyst is looking for. In addition, the level of detail may be varied, or the full details may be presented in another view.

1) Cost of knowledge: An important measure in designing an effective interaction scheme is the *cost of knowledge*: The amount of energy that must be invested to extract a certain amount of information. Pirolli and Card compare a human's search for information to an animal's foraging for food; both seeking to minimize the amount of energy required to cover their needs. It is noted that information often appears in chunks or clusters separated by long distances of uninteresting data, just like edible plants do in nature. Furthermore, as scents may direct an animal to the food source, there may be hints in a visualization directing navigation towards the more interesting information [17].

This leads to a design methodology of analysing and minimising the effort needed to extract knowledge from a visualization system. The cost of knowledge is twofold: First, there is the actual effort of locating and extracting the information. Then, one also has to consider the cost of not being able to do something else during the information search [1].

To minimize the cognitive effort, two aspects should be addressed:

- Reducing the cost of analysing the visual representation: Improving the visualization to better support preattentive processing, and increasing compliance with cognitive limitations.
- Reducing the cost of navigating the visualization by choosing efficient interaction techniques.

2) A cognitive interaction model: The process of working with an interactive visualization can be described as three nested loops that are being performed by the user [1]: The problem-solving loop, the exploration and navigation loop, and the data manipulation loop. On the highest level, the analyst forms hypotheses about the data and refines these through revision of the visualization. This is supported by the exploration and navigation loop, where the visualization is navigated and the user builds a mental "map" of the data that is presented. On the lowest level, data items are identified and selected using basic motoric actions (such as eye-hand coordination).

In order to support the flow of the user's work, the system must be responsive enough to avoid interrupting the thinking process. Shneiderman states that response times for mouse and typing actions should be less than 0.1 second [18]. While lower response times generally lead to higher user satisfaction, there is a danger that the fast pace will increase the user's error rate.

Another aspect of response time is related to what is known as *change blindness*: If the user blinks or changes focus, or if the screen briefly goes blank, between two successive views, it is probable that the user will be unable to detect the difference between the views. Even when searching actively for a difference, this task is difficult [11]. This means that the visualization should not presume that a user is able to detect changes between views without a way of highlighting the change, such as with animation.

3) Interaction techniques: Given the cognitive interaction model described above, it must be a goal to find techniques supporting the two inner loops: Explore/navigate and data manipulation. This calls for effective navigation techniques using simple manipulations to find the information of interest, always supporting the higher-order goal of finding, refining or investigating a hypothesis. Ben Shneiderman describes what he calls the Visual Information Seeking Mantra: "Overview first, zoom and filter, then details-on-demand" [19], to which he devotes ten repeating lines in his article – once for each project in which it was rediscovered.

This indicates the need for a visualization that serves different purposes at different stages in the user's process of gaining understanding of the data: First, an overview needs to be presented, to give the user a clue about what to look for. Then, relevant information must be sorted out, and all details should be easily accessible when the user needs them.

It is interesting to note Colin Ware's [1] comments to this approach: He suggests that the process is not as directed as Shneiderman claims it to be, but rather an iterative procedure where interesting features are spotted, the view is zoomed out to get an overview, and then zoomed back in again for inspection of the details. Ware concludes that no matter how the process turns out to be performed, it is highly important that the visualization acts as an interface capable of performing these operations.

Shneiderman provides a taxonomy of important visualization and interaction tasks [19]. While the first four tasks (overview, zoom, filter, details-on-demand) are frequently found in prototypes, the three latter (relationships, history, extraction) are more seldom since they do not add to the capability of the visualization itself, "only" to the user interface. They are, however, important for the explorative aspects of interaction and should be remembered when developing a prototype.

D. Examples of related visualizations

The research on health information visualization within health informatics is dominated by few, but active groups. A good review of EHR-visualization from these groups is given by Alexander Rind et. al. [20]. However, the information visualization research field is much wider, and many relevant methods and techniques are not health-specific. We have limited the review to approaches that had impact our design.

1) LifeLines[3], [4], [5]: A LifeLine is an interactive timeline visualisation, plotting events for a single history grouped into different *facets*. Each facet contains information that is semantically related, such as (in a health record setting) diagnoses, medications, allergies, and general problems (e.g. smoker, depression). The visualisation can show information at different levels of abstraction: For example, medications can be shown using a name for the group of drugs (beta blocker) or by the individual drug names (athenolol, propanolol). The prototype supports zooming, and searching for related items (i.e. searching for "migraine" highlights all diagnoses and drugs related to migraine). In addition, attributes can be mapped to different graphical representations by the user.

2) Other Examples: Another direction that is relevant to our work is temporal queries and the results of such queries. Chittaro and Combi [21] describe several metaphors for describing intervals with uncertain length: An elastic band, a spring, or a strip of paint. Representations of physical objects constrain the length of these representations to appeal to the user's real-world experience. Their studies include usability tests of the different representations.

The visualisation used by Fails et al. can remind of an event chart showing multiple lines per history, one for each hit of a temporal query [22]. However, the visualisation shows only the time spanned by the search hits, as opposed to the traditional event chart showing the entire histories. Also, events not part of a search hits are only counted in the design of Fails et al., while an event chart typically treats all events selected for display equally. Concerning the relation between event charts and life course visualisations, both show history information in a timeline view. While event charts (and the design of Fails et al.) show many histories in the same view, the life course visualisations show much more information for the one history they show.

The Clinical Narrative Temporal Reasoning Framework (CNTRO) [23] is interesting, because it designed to capture, represent and reason with the temporal semantics of events, intervals and their constraints in EHR. In retrospect, we have implemented much of the same functionality, but in a less well-founded way. Currently, we are investigating the use of constraint logic programming to handle interval reasoning.

III. Case to be investigated

We choose to concentrate on chronically ill patients as they frequently have complex patient histories. The data we have available for this research is from another research project that investigates patients trajectories in a prospective longitudinal cohort study with data on somatic primary and specialist health care utilization for a two-year period. Visits to a somatic health care service includes any visit to a hospital (inpatient, outpatient or day treatment), receiving services from the adjacent municipalities (home care services, nursing home etc.) and visits to a primary care provider (General Practitioner (GP), emergency primary care services operated by GPs, physiotherapist etc.) or private medical specialist where the provider had claimed reimbursement. The imported data was structured, i.e. the data is available in given fields and coded in a standard way. For example, diagnoses are mainly coded using ICPC-2 and/or ICD-10.

IV. Results - Design and Use

The prototype was used in the research project to select 13,000 patients from a data set of 168,000 patients based on predefined characteristics. For the 13,000, their individual trajectories was created using the prototype and presented to the patients in a simplified form[3] to get their feedback on their experiences with their own trajectories. In our experiment, only 1% of the patients said that everything was wrong in the presented trajectories/contacts we thought they had had with the health service, while 92% could easily recognize their own trajectory and 7% did not remember [24].

With basis in and references to the work described in the preceding section, this chapter proposes a visualization with accompanying interactive features. The proposed visualization is implemented as a Java prototype; a screenshot of this prototype is shown along with a description of its user interface in Figure 4. The visualization shows each patient history as a bar annotated with symbols representing the events in the history, and interval concepts shown by background colorings. Interactive operations on this diagram include extraction of sub-collections, sorting and aligning histories, filtering events, and searching for temporal patterns.

To speed up drawing and to become more independent of the database schema, all content to be visualized or queried is pre-loaded into a data structure of Java objects. The entries themselves are either intervals, defined by their start and end times, or events that happen at a given time and have no duration. Intervals could be notions such as Hospital stay. Concerning point events, these are single day contacts, usually with a recorded diagnosis. When it comes to the representation of time, entries with a clearly invalid date (prior to the birth of the patient) are ignored.

A. Regular Expressions

We use regular expressions to describe subsets of the above hierarchies. The main motivation for this is simplicity: Regular expressions are readily supported by our programming environment (Java), and with a regular expression one may easily refer to any branch of the hierarchies by listing the first few letters or digits and appending a wildcard. Such specifications may be combined using the disjunctive construct of regular expressions; so to specify diagnoses concerning the eye (F) or ear (H) one may specify the regular expression: `F.*|H.*`. While being a useful tool for computer scientists, general practitioners cannot be expected to be acquainted with regular expressions. This means that a graphical user interface is needed (see Figure 4). Regular expressions are also used for extraction of some of the available free text data, as explained below. However, this extraction is limited because of differing conventions and many typing errors in the text.

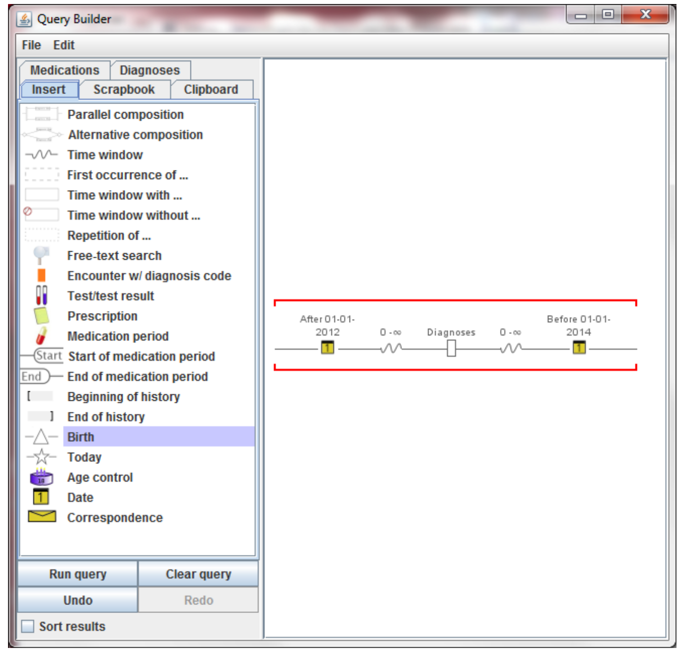


Fig. 4. Query-Builder Interface

B. Axes

To make it possible to address individual patients, patient ID numbers (taken from the database) are shown along the vertical axis. The horizontal axis has two modes:

1) When the diagram is not aligned, the axis shows calendar time (the actual dates).
2) In an aligned diagram, the axis shows the number of months before and after the alignment point.

In order to accommodate both the many short trajectories, and the few long ones, two sliders were added to the user interface (see bottom right of Figure 1). The sliders allow the user to zoom both vertically and horizontally, in order to see many patients and/or many details (long time-span) at the same time.

V. Conclusion

To learn more about groups of patients and their treatment processes clinicians and patients can be greatly helped by getting the information visualized. In addition, we believe that the visualization can be useful to researchers looking at data to be statistically evaluated, in order to discover new hypotheses or get ideas for the best analysis strategies. We think that the tool we have developed is usable for reasonable large data set, but it can be challenging to use for very large data sets.

Acknowledgment

The authors would like to thank all present and previous members of the PAsTAs project for support and data collection, and for guidance, advice, and for interesting discussions, in particular Ole Edsberg and general practitioner and professor Anders Grimsmo.

[3] see http://pastas.no, sample password: tromsø

References

[1] C. Ware, *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers, 2004.

[2] M. D. Maltz and J. M. Mullany, "Visualizing lives: New pathways for analyzing life course trajectories," *Journal of Quantitative Criminology*, vol. 16, no. 2, pp. 255–281, 2000.

[3] C. Plaisant, D. Heller, J. Li, B. Shneiderman, R. Mushlin, and J. Karat, "Visualizing medical records with lifelines," in *Proceedings of ACM CHI 98 Conference on Human Factors in Computing Systems (Summary)*, ser. Demonstrations: Interactive Medicine, vol. 2, 1998, pp. 28–29. [Online]. Available: http://www.acm.org/pubs/articles/proceedings/chi/286498/p28-plaisant/p28-plaisant.pdf

[4] A. Snyder, B. Shneiderman, C. Plaisant, D. Heller, J. Li, K. P. Colorado, and R. Mushlin, "Lifelines: Using visualization to enhance navigation and analysis of patient records," Nov. 18 1998. [Online]. Available: http://citeseer.ist.psu.edu/32042.html; ftp://ftp.cs.umd.edu/pub/papers/papers/ncstrl.umcp/CS-TR-3943/CS-TR-3943.ps.Z

[5] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman, "Lifelines: Visualizing personal histories," Maryland University, Tech. Rep., Oct. 15 1995. [Online]. Available: http://hdl.handle.net/1903/440

[6] Ø. Hauge and S. J. Nordbø, "Nsepter implementation documentation," Norwegian Centre of Electronic Patient Records Research, Tech. Rep., 2005.

[7] S. J. Nordbø, "Visualising collections of event sequences from general practice patient records," NTNU, Tech. Rep., 2005.

[8] B.-S. Seah, S. S. Bhowmick, C. F. Dewey, Jr, and H. Yu, "FUSE: Towards Multi-level Functional Summarization of Protein Interaction Networks," in *Proceedings of the 2Nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, ser. BCB '11. New York, NY, USA: ACM, 2011, pp. 2–11. [Online]. Available: http://doi.acm.org/10.1145/2147805.2147806

[9] C. Jin, S. S. Bhowmick, X. Xiao, J. Cheng, and B. Choi, "GBLENDER: Towards Blending Visual Query Formulation and Query Processing in Graph Databases," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '10. New York, NY, USA: ACM, 2010, pp. 111–122. [Online]. Available: http://doi.acm.org/10.1145/1807167.1807182

[10] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Addison-Wesley, 2005. [Online]. Available: http://www-users.cs.umn.edu/ kumar/dmbook/

[11] D. J. Simons and M. S. Ambinder, "Change blindness. theory and consequences," *Current Directions in Psychological Science*, vol. 14, no. 1, pp. 44–48, 2005. [Online]. Available: http://www.blackwell-synergy.com/doi/abs/10.1111/j.0963-7214.2005.00332.x

[12] C. G. Healey, "Fundamental issues of visual perception for effective image generation," in *SIGGRAPH 99 Course 6*, 1999.

[13] S. Faisal, A. Blandford, and H. W. Potts, "Making sense of personal health information: challenges for information visualization," *Health informatics journal*, vol. 19, no. 3, pp. 198–217, 2013.

[14] V. L. West, D. Borland, and W. E. Hammond, "Innovative information visualization of electronic health record data: a systematic review," *Journal of the American Medical Informatics Association*, vol. 22, no. 2, pp. 330–339, 2015.

[15] G. Kopanitsa, C. Hildebrand, J. Stausberg, K. Englmeier *et al.*, "Visualization of medical data based on EHR standards," *Methods Inf Med*, vol. 52, no. 1, pp. 43–50, 2013.

[16] Y. Zhang, R. Padman, and N. Patel, "Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data," *Journal of biomedical informatics*, vol. 58, pp. 186–197, 2015.

[17] P. Pirolli and S. K. Card, "Information foraging in information access environments," in *CHI*, 1995, pp. 51–58.

[18] B. Shneiderman, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 3rd ed. Addison Wesley Longman, 1998.

[19] ——, "The eyes have it: A task by data type taxonomy for information visualizations," Sep. 05 1996. [Online]. Available: http://citeseer.ist.psu.edu/175307.html; ftp://ftp.cs.umd.edu/pub/papers/papers/ncstrl.umcp/CS-TR-3665/CS-TR-3665.ps.Z

[20] A. Rind, T. D. Wang, A. Wolfgang, S. Miksch, K. Wongsuphasawat, C. Plaisant, and B. Shneiderman, "Interactive information visualization to explore and query electronic health records," *Foundations and Trends in Human-Computer Interaction*, vol. 5, no. 3, pp. 207–298, 2011.

[21] L. Chittaro and C. Combi, "Representation of temporal intervals and relations: Information visualization aspects and their evaluation," in *TIME*, 2001, pp. 13–20. [Online]. Available: http://www.computer.org/proceedings/time/1107/11070013abs.htm

[22] J. A. Fails, A. Karlson, L. Shahamat, and B. Shneiderman, "A visual interface for multivariate temporal data: Finding patterns of events across multiple histories," in *Visual Analytics Science And Technology, 2006 IEEE Symposium On*. IEEE, 2006, pp. 167–174.

[23] C. Tao, K. Wongsuphasawat, K. Clark, C. Plaisant, B. Shneiderman, and C. G. Chute, "Towards event sequence representation, reasoning and visualization for ehr data," in *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. ACM, 2012, pp. 801–806.

[24] H. D. Wågbø, "The patient perspective: Utilizing information visualization to present health records to patients," p. 131, 2014.