Visual Clustering Analysis of CIS Logs to Inform Creation of a User-configurable Web CIS Interface

Y. Senathirajah[1]; S. Bakken[1, 2]
[1]Department of Biomedical Informatics, Columbia University, New York, NY, USA;
[2]School of Nursing, Columbia University, New York, NY, USA

Keywords

Clinician information needs, widget-based interface, clinical information system interface, web2.0

Summary

Background: In this paper, we describe a new method for the study of clinical information system (CIS) logfiles joined with information in the clinical data warehouse. This method uses heatmap representations and clustering techniques to examine clinicians' viewing patterns of laboratory test results. The context of our application of these techniques is to inform the creation of a widget-based interface to the CIS.

Objectives: We address the rationale, feasibility, and usefulness of our method through examination of three hypotheses:

1) The frequency distribution of laboratory test viewing will follow a 'long tail' pattern, indicating that patterns are highly variable and supporting the rationale for a widget-based configurable system.

2) Patterns of laboratory testing viewing (by clinician, specialty, clinician/patient/day, and ICD-9-CM codes) can be distinguished by our methods.

3) The identified clusters will include more than 80% of the laboratory test elements found in 30 randomly selected patient records for one day.

Methods: The data were plotted as heatmaps and clustered using hierarchical clustering software. Various parameters were tested to give the optimal clusters.

Results: All the hypotheses were supported. For Hypothesis 3, 91.4% of information elements in the records were covered by the generated clusters.

Conclusions: Study findings support the rationale, feasibility, and usefulness of our methods to examine patterns of information access among clinicians and to inform the creation of widget-based interfaces. The results also contribute to our general understanding of clinicians' CIS use.

views by selecting and arranging (dragging and dropping) elements of the CIS according to their own requirements. Similar interfaces such as iGoogle are found in the public Web. A CIS incorporating such an interface might be expected to have many advantages, including:

- greater suitability to user needs reflecting users' medical domain knowledge,
- efficiencies due to workflow and other human-computer interface (HCI) improvements,
- sharing of user-created components,
- capture of users' tacit domain and institutional knowledge,
- incorporation of diverse information sources,
- agile reconfiguration to meet emerging health challenges [3], and
- improved user acceptance and adoption, communication, and collaboration [4].

Despite these expected advantages, widget-based interfaces are not likely to be useful unless widgets are pre-populated with relevant data elements based upon clinically meaningful clusters. Thus, we applied heatmaps and clustering techniques in the domain of clinician laboratory test viewing: 1) to confirm the need for user-configurable interfaces, such as MedWISE, through examining the distribution of clinician viewing patterns for laboratory results; 2) to determine if clinician viewing patterns for laboratory results can be clustered in meaningful ways to inform the design of widget collections that may meet targeted purposes (e.g., renal transplant) or more general purposes; and 3) to ascertain if the clusters of laboratory results generated through our method capture, to a useful degree, the laboratory results in a commercial electronic health record (EHR) for

1. Introduction

In this paper, we describe a new method for the study of clinical information system (CIS) logfiles. The method uses heatmap representations and clustering techniques for examining clinician viewing patterns. This is in contrast to prior methods, which typically apply more conventional statistical approaches.

Although the new method may be useful for other aspects of CIS design, the context of our application of these techniques was to inform the creation of a widget-based CIS interface, MedWISE [1, 2]. MedWISE allows users to create their own tools and

a small random sample of patients. The last is necessary to determine the proportion of data that would not be included in a widget-based interface that was pre-populated based upon the clusters; if the clusters do not include a substantial portion of patient data from actual clinical records, little time would be saved by pre-populating interfaces based upon the clusters. In this initial report, we address the rationale, feasibility, and usefulness of the method of applying heatmaps and clustering techniques.

2. Background

As foundation for our study hypotheses and methods, we delineate the relevance of analyzing clinician viewing patterns to inform the design of widget collections for interfaces such as MedWISE, describe the CIS used in our analysis, and provide a brief background on microarray analysis and clustering techniques.

2.1 Analyzing Clinician Viewing Patterns to Inform Design of Widget Collections

Understanding the composition of a clinically useful set of widgets is vital to the success of introducing a widget-based CIS user interface. Logfiles offer a granular and detailed picture of what clinicians actually view in the CIS, and thus logfile analysis has the potential to inform the creation of suitable widget collections. For example, Chen and colleagues [5, 6] described a large variety of patterns of information viewing in a logfile analysis of temporal patterns of information access in WebCIS, the CIS in our study. They found that while 30–40% of data elements viewed were common across clinicians, the rest differed, requiring the clinicians to take different paths (which varied with specialty, role, patient condition, and other factors) through the system. As an indication of how variable these patterns are, Chen et al. found each specialty used 300–500 different paths.

Part of the rationale for a configurable widget-based interface is that it will be able to accommodate a high degree of variation in information viewing, i.e., cater to the

'long tail'[a] [7–9], and thus more easily suit different users. Accordingly, we were interested in discovering whether or not the laboratory results viewing frequency distribution in our sample follow a long tail pattern (>50% of views taking place in the 'tail' part of the distribution curve). If a majority of laboratory tests viewed by clinicians represented only a few types of tests that were frequently viewed, traditional methods of modeling a small number of use cases and creating static software displays would be adequate. Even in the absence of a long-tailed frequency distribution, clusters of viewing patterns can inform the pre-population of the CIS interface with a suitable collection of widgets to meet user needs.

2.2 CIS Details

Several aspects of our CIS provide the foundation for our analysis. WebCIS is a CIS used by clinicians throughout the Columbia campus of New York-Presbyterian Hospital (NYPH) [10]. It provides laboratory results, clinician documentation notes, imaging reports, forms for note entry, demographic and billing data, and listings of other available data. WebCIS functions as a Web portal-like interface, integrating many NYPH information systems, both by providing links to independent systems and by providing clinical information from multiple systems via the clinical data repository (CDR), a cache-like integrating database. The same data, in a form optimized for research queries, is also stored in the clinical data warehouse (CDW). A standardized terminology, the Medical Entities Dictionary (MED)[10, 11], provides integration and translation of terminologies between different systems and assigns each term a unique non-semantic identifier, a medcode. Thus, data el-

---

[a]   A 'long tail' distribution is one in which there is a large variety of different objects that each occur at low frequency, as opposed to a collection that has a few objects, each occurring at high frequencies. An example would be academic books (many different ones are each sold in small numbers) versus bestsellers (a few different books are each sold in large numbers). For more information on the 'long tail' concept, please see 'The Long Tail' by Chris Anderson, Wired, April 2004. http://www.wired.com/wired/archive/12.10/tail.html.

ements may be referred to at a granular level by their medcodes, which are included in, or can be derived from, logfiles. Because WebCIS forces a new server request for each laboratory test or panel, logfiles provide very granular information about what specific laboratory results are viewed by which clinicians and for which patients.

2.3 Microarray Analysis

Analysis of CIS data at the granularity needed to inform widget interface creation presents a problem in multidimensional analysis and interpretation because there are hundreds of laboratory tests whose results are viewed by multiple clinicians for multiple patients. For our purposes, it is important to identify data elements commonly viewed by most clinicians and by subgroups of clinicians, as well as those only viewed by individuals or that are associated with a certain specialty or patient condition. An understanding of all these, as well as of how they may be clustered into patterns, is necessary to design the initial sets of widgets suitable to the specific user's needs and situation, organize them into a widget-based CIS interface, and enable predictive delivery as the user traverses the CIS. Prior work in related areas includes the study of Wright and Sittig [12] in using the apriori algorithm for frequent item set mining and association rule mining to identify order sets in an automated fashion; however, we are aware of no others using our clustering techniques for automated data presentation in CISs. Wright and Sittig point out that automated methods which are used in a system which already has encoded the relevant elements and terminologies for the target CIS are extremely economical, eliminating several of the time-consuming steps which would otherwise be required, and reflecting local practice. One difference between the aims of their study and ours is that they were concerned only with frequent sets; we are concerned with those that may be infrequent as well. Consequently, we applied an alternative approach built upon techniques usually used in microarray analysis.

Microarrays and clustering algorithms have been used for gene expression analyses
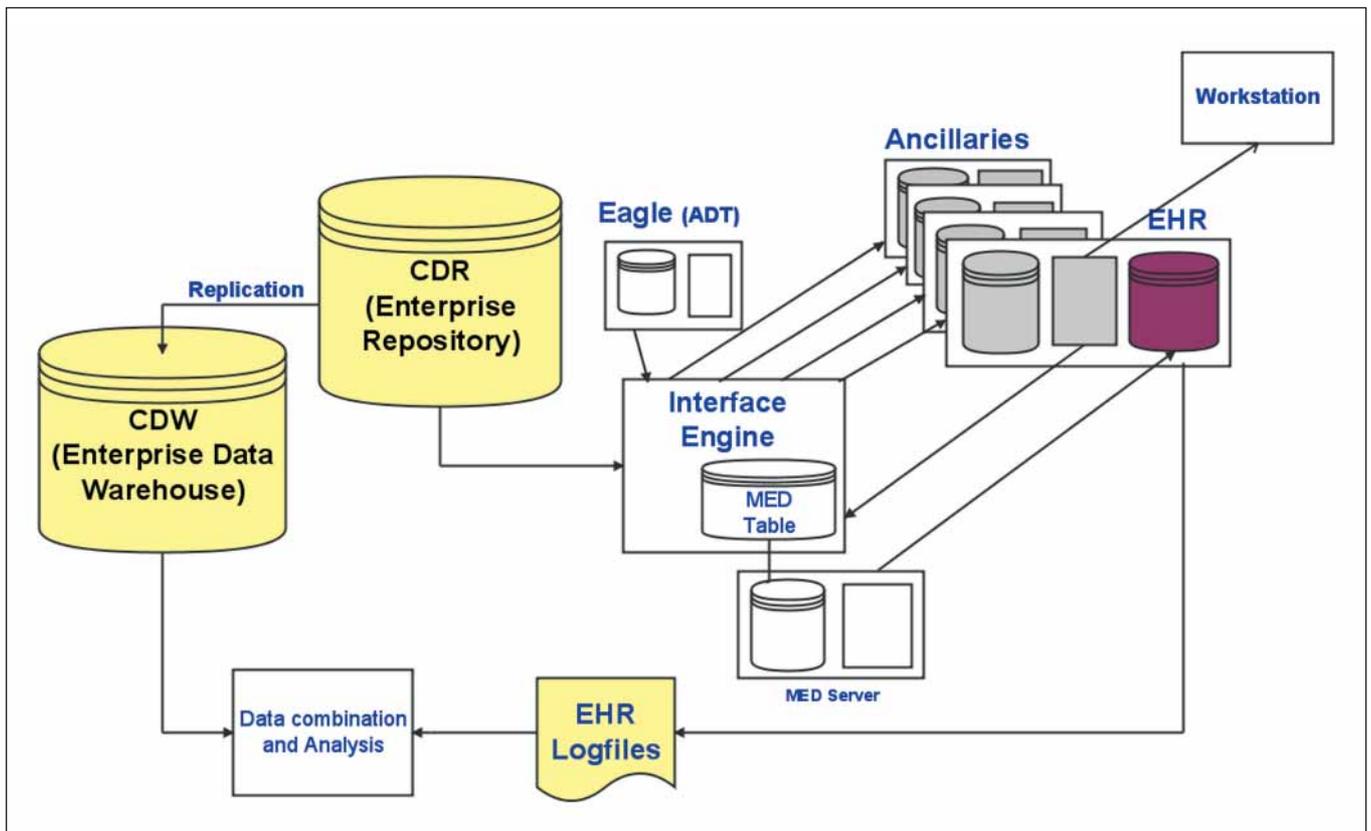
Fig. 1 System architecture as it pertains to logfile and CDW information. The system consists of many commercial and home-grown systems denoted by 'CIS' and 'Ancillaries,' which, together with the Admission, Discharge, and Transfer (ADT) system (Eagle), use a series of interfaces (Interface Engine) to transfer information between systems and store and retrieve data from the CDR, a database system optimized for speed. The interface engine draws on the MED server for terminology translation between the different systems. Data from the CDR is replicated to the CDW after processing by scripts, which put it into a relational form suitable for research queries. Entries in CDW tables include medcodes for each laboratory test or other clinical elements. The CIS generates logfiles that contain either a medcode or unique time-stamp, which can be joined with data in the CDW to identify elements viewed.

of hundreds of samples and factors. [13–15] Madeira and others [15–18] list many possible types of biological information which can be derived from clustering analyses of microarrays, including gene functional annotation, identification of genes which are co-regulated/co-expressed, classification of samples, and so on. Heatmaps have also been used to display clinical chemistry data, [19] business processes [20], and steroid hormones in metabolism studies [21]. For our purposes, the task of identifying useful combinations of CIS data elements is most similar to that of identifying genes which are co-expressed under particular conditions. We conceptualize the clinician (or patient condition, or user session) as analogous to a 'sample' and the array of laboratory tests as analogous to genes. Consequently, we consider the frequency of viewing laboratory results

as analogous to gene expression. In our case, each record represents a clinician's total scores for viewing each laboratory test. Thus, for each of the 30 clinicians and 598 laboratory panels, we have a count of the total number of times the clinician viewed the laboratory results. As in genomic studies, these data can be plotted in a heatmap, which is like a spreadsheet with colors representing ranges of values in each cell (▶ Fig. 2). By convention, shades of red indicate high values, green low values, and black intermediate values.

Heatmaps from microarrays have two useful features of particular relevance to our study; they can display clusters in an intuitive way, and the visualization of numbers (as colors) associated with the rows and columns (laboratory test and clinician, diagnosis, specialty, etc.) is itself a useful visualization of information use pat-

terns, even without clustering. For example, a nonclustered heatmap showing elements along the x-axis and individual clinicians' use (grouped by specialty) along the y-axis would show any differences in viewing patterns among individual clinicians and clinician specialties.

2.4 Clustering Techniques

Clustering rearranges the heatmap to place records with similar viewing patterns together, forming contiguous areas with similar color. Dendrograms (tree-diagrams) along two axes show the relationships of the rows and columns, with the most similar elements placed next to each other and joined by lines of the tree (as with a phylogenetic tree). Similarity is based on a distance measure–a multidimensional

analog to two-dimensional Euclidean distance. For more information, see [22, 23].

Kriegel et al. [13, 14] categorize clustering methods into three categories: subspace clustering, pattern-based clustering, and correlation clustering, and state that each defines the problems it is trying to solve differently, making comparison difficult. Pattern-based clustering methods identify simple correlations, whereas subspace clustering techniques are used for more complex numerical data in which magnitudes are significant. Correlation clustering methods identify more complex correlations in which there are complicated functions relating the variables. We selected the hierarchical (pattern-based) clustering methods usually used in microarray analysis because of the nature of the problem we are trying to solve, which is to find the useful patterns in clinician data viewing in order to provide useful widgets and clusters of widgets in the CIS interface. Advantages of hierarchical clustering algorithms are that investigators do not need to specify the number of clusters in advance and results are presented visually in the form of a dendrogram and heatmap. The algorithm is unsupervised, meaning clusters emerge solely from the data and do not require a 'gold standard' set of labeled examples for comparison and learning.

One difference between the tasks in our study and that of most biological analyses is that the latter analyze expression patterns in order to elucidate facts about putative underlying biological structures or relationships, such as co-expression of genes indicating that they are co-regulated and perhaps part of the same pathway fragment. The patterns are noisy and imperfect clues to the underlying sought knowledge. In our case, the co-occurrence patterns are themselves what we are seeking, rather than being imperfect pointers to a behind-the-scenes process. In addition, distance between clusters may also not have the same significance it would in microarray analysis of biological data. For example, tests for sexually transmitted diseases may be found
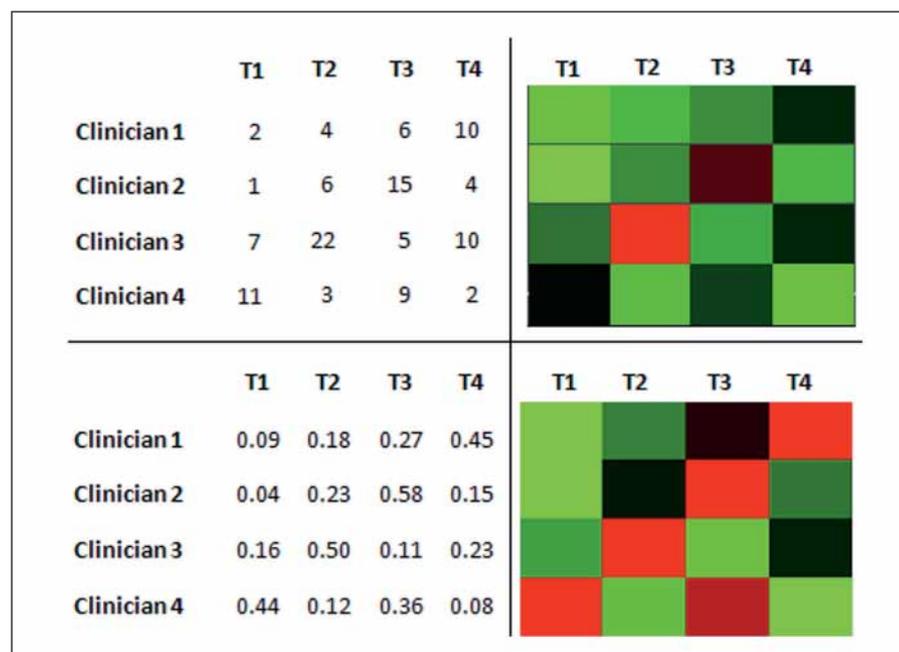
Fig. 2 Cell values correspond to colors in a heatmap. By convention, shades of red indicate high values, green low values, and black intermediate values. Making an analogy with microarrays in our study, laboratory tests are analogous to genes, and clinicians (or patient condition, or user session) are analogous to samples. The top row shows raw values and the corresponding heatmap. The bottom row shows the values normalized by row, so that each value is expressed as a a difference of the value from the row mean, over the row standard deviation. (The row constitutes the total views for that clinician). The corresponding heatmap is to the right.
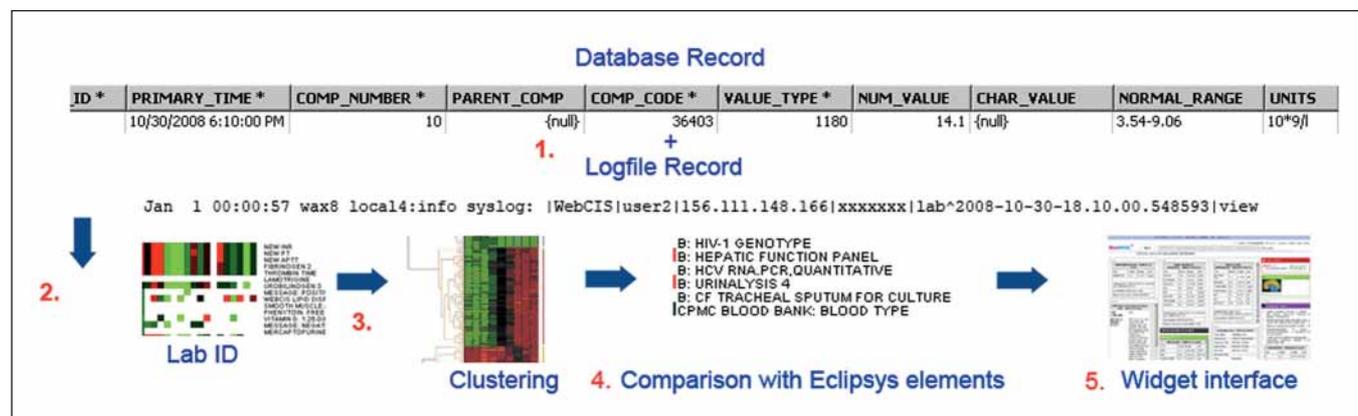
Fig. 3 Methods Overview: 1) Data from the database record and logfile record are joined to allow identification of which element (in this case, specific laboratory results) the user viewed for a specific time and patient. 2) Normalized counts of all the test results viewed are visualized as a heat map, and clustered hierarchically in two dimensions (e.g. laboratory test v. user). 3) Cluster analysis is performed. 4) Resulting clusters are compared with actual patient records from the Eclipsys CIS (a commercial system in use at NYPH). 5) The cluster can be used to specify elements of a widget interface (future work).

together with tests of renal function even in the absence of a biomedical linkage. While it is true, and highly interesting, that patterns of information viewing in a CIS may correspond to medical and environmental facts, or facets of the diagnostic and treatment processes, elucidating such processes in detail is beyond the scope of our current work. We are not seeking specific patterns, but to find the statistical regularities corresponding to clinicians' prototypical viewing patterns in order to customize CIS display in a way responsive to clinicians' needs. Knowledge-based approaches can be combined with this determination of clusters, as Mobasher et al. describe in their studies [24, 25] of commercial web mining and personalization. Commercial web mining and personalization techniques make use of page content analysis and/or user tagging or ratings in combination with usage analysis. CIS information does not have direct analogues to page content, but there may be other sources of knowledge to shape personalization; the MED hierarchy is an example. However, while we may use a combination approach in future, this is not within the scope of the work described here. Examining co-occurrence patterns from logfiles allows us to determine which elements clinicians usually view in the same session, making those combinations good candidates for being displayed together on the page in pre-populated interfaces. Usual methods of determining interface elements include consultation with clinical users in an iterative participatory design approach that is usually discontinued at some point during development or implementation. Other approaches include cognitive studies of system use [26–28], which, due to cost, usually involve small numbers of subjects over a limited time. However, because of the configurable nature of MedWISE, we required a method that would allow us to study clinician viewing patterns at a granular level.

2.5 Objectives

To understand how clinicians view specific types of information at a very fine-grained level to inform design of widget-based interfaces, we examined the rationale, feasibility, and usefulness of applying heatmaps and hierarchical clustering through three hypotheses:

- Hypothesis 1: The frequency distribution of laboratory results viewing will follow a 'long tail' pattern.
- Hypothesis 2: Patterns of laboratory testing viewing (by clinician, specialty, clinician/patient/day, and ICD-9-CM codes) can be distinguished by our methods.
- Hypothesis 3: The identified clusters will capture 80% of the laboratory results data elements in a random sample of 30 patient records for one day.

3. Methods

In this section, we first describe the preparation of log files for cluster analysis and then the analysis techniques. ▶Figure 3 provides an overview of methods starting with the database record and resulting in the widget-based interface. The last step in the overview is future work and not the focus of this paper, but is provided as context for the remaining methods.

3.1 Preparation of Log Files for Cluster Analysis

We extracted one month (January) of log entries for 30 clinicians in the Nephrology and Hospitalist departments from 2007 WebCIS logfiles. After pre-processing to remove redundant or irrelevant information (such as logfile entries indicating an error that prevented information viewing, or entries pertaining to viewing of non-laboratory data), we processed the logfiles with data from the CDW as described above to identify all laboratory results viewed by clinicians in our sample. For this study, we considered the unit of analysis to be a view of a laboratory test result panel (a small group of tests, such as those included in standard basic metabolic panels) or a single test when displayed alone rather than as part of a panel. This was done to remove the bias that would occur from using each individual test in a panel as the unit of analysis since individual tests in a panel would

naturally show up as clusters in our visualizations and this clustering merely reflects design of the laboratory ordering schemas and provides no useful information about clinician viewing preferences. The resulting datasets were manipulated in a database and spreadsheet to execute various joins and lookups necessary to isolate the laboratory records and to put the data in a format usable by the clustering software.

3.2 Cluster Analysis

For analysis we used the open-source freely downloadable Hierarchical Clustering Explorer available from the University of Maryland [29] and Spotfire for Functional Genomics version 9.1.1. These programs perform hierarchical, K-means, and other clustering techniques, and provide visual interactive displays for profile discovery.

The sample comprised logfiles from 30 clinicians pertaining to 1239 patients; 653 laboratory test medcodes were represented. Analysis using laboratory panels as a single unit of analysis reduced the number of medcodes to 598.

After assessing various distance measures (Euclidean, simple matching, Manhattan, etc.) and linkage types, we clustered the data according to the analyses below. In general, we used complete linkage and Pearson's r as a distance measure, because of its independence of units and because it generally produced the best results in terms of consistent and well-defined useful clusters.

3.3 Hypotheses Testing

3.3.1 Hypothesis 1

We plotted viewing frequency for each laboratory element and used descriptive statistics to determine whether >50% of the views were for the less common tests (the 'long tail'). We also plotted the cumulative percentage of test views on the same x-axis to show the proportion of test views corresponding to different parts of the frequency distribution.

Fig. 4 Distribution plot of test viewing frequency (vertical axis) versus laboratory tests (horizontal axis, not all test names shown). The distribution shows a 'long tail' pattern. The darker line plot is the cumulative percentage (Pareto plot), showing that >50% of total test views are of tests in the tail of the frequency distribution.

3.3.2 Hypothesis 2

We analyzed viewing patterns by clinician, specialty, day (views by clinician per patient per day), and ICD-9-CM codes (which are included in the CDW as part of the patient record). Data were clustered in Spotfire using the correlation similarity measure, complete linkage. We also analyzed the comparative viewing patterns of matched pairs of users who viewed the same patient record on the same day. This included the calculation of overlap in test viewing between members of the pair and calculation of the proportion of sessions in which test viewing was completely mutually exclusive between members of the pair. The Jaccard index (or Jaccard similarity coefficient) is a statistic used for comparing the similarity and diversity of sample sets. It consists of the intersection of the two sets divided by their union.

3.3.3 Hypothesis 3

As a test of the usefulness of the clustering to inform the design of the widget-based interface, we first manually inspected the records of 30 nephrology patients (1% of the number of patients represented in log-

files) selected at random for one inpatient day, counting the number of laboratory tests viewable in Eclipsys (a commercial CIS in use at NYPH). Next, we examined differences between these and the cluster-determined laboratory sets. We defined 'usefulness' as how well clusters found through our method represent elements found in real patient records, i.e. we calculated the proportion of laboratory elements found in real patient records which were also found in the clusters we found from logfile analysis. The 80% threshold used for this hypothesis is somewhat arbitrary and loosely based on the 80/20 rule found in so many work analyses(30).

4. Results

4.1 Hypothesis 1

The test viewing frequency analysis revealed that 51.8% of the lab panel views were of 498 different tests; 49.2% of the views were of the top 10 tests. 10 tests corresponds roughly to the number of widgets that can be placed in a single-view interface, and so reflects a natural breaking point for our purposes. It also reflects the nature of the

distribution which consists of a a small number of high-frequency tests and a much larger number of lower-frequency tests. The number of views ranged from 5527 (for basic metabolic panel) to 1 (for many specific disease or single-test items). This distribution supports the 'long tail' characteristic of medical information access (►Fig. 4).

4.2 Hypothesis 2

Hierarchical clustering on both the med-code values (columns) and clinicians using a minimum similarity of .85 (complete linkage) yielded several clusters (►Fig. 5). These included clinically-meaningful clusters related to bone health (Vitamin D, parathyroid hormone), transfusion (anti-erythrocyte antibody screen, blood type), diarrhea of possible infectious origin (parasitology exam, fecal leukocytes, stool culture, Clostridium difficile Toxin A), and autoimmune disease (rheumatoid factor, complement, total hemolytic (ch50) complement, c4 complement, c3 complement, anti nuclear antibody screen).

Fig. 5 Heatmaps with laboratory test titles (vertical axis) and clinicians (horizontal axis, names obscured). On the right is an overview screenshot of the heatmap showing clusters of tests and user patterns. Horizontal groupings reflect tests commonly viewed together. Vertical bands reflect individual clinician viewing patterns. On the left is an exploded view of a section of heatmap showing the test names on the vertical axis. Horizontal red bands indicate the tests that were viewed by many users and at high frequencies. Conventional heatmap coloring (i.e., red for high numbers, green for the lowest, black for intermediate) applies.
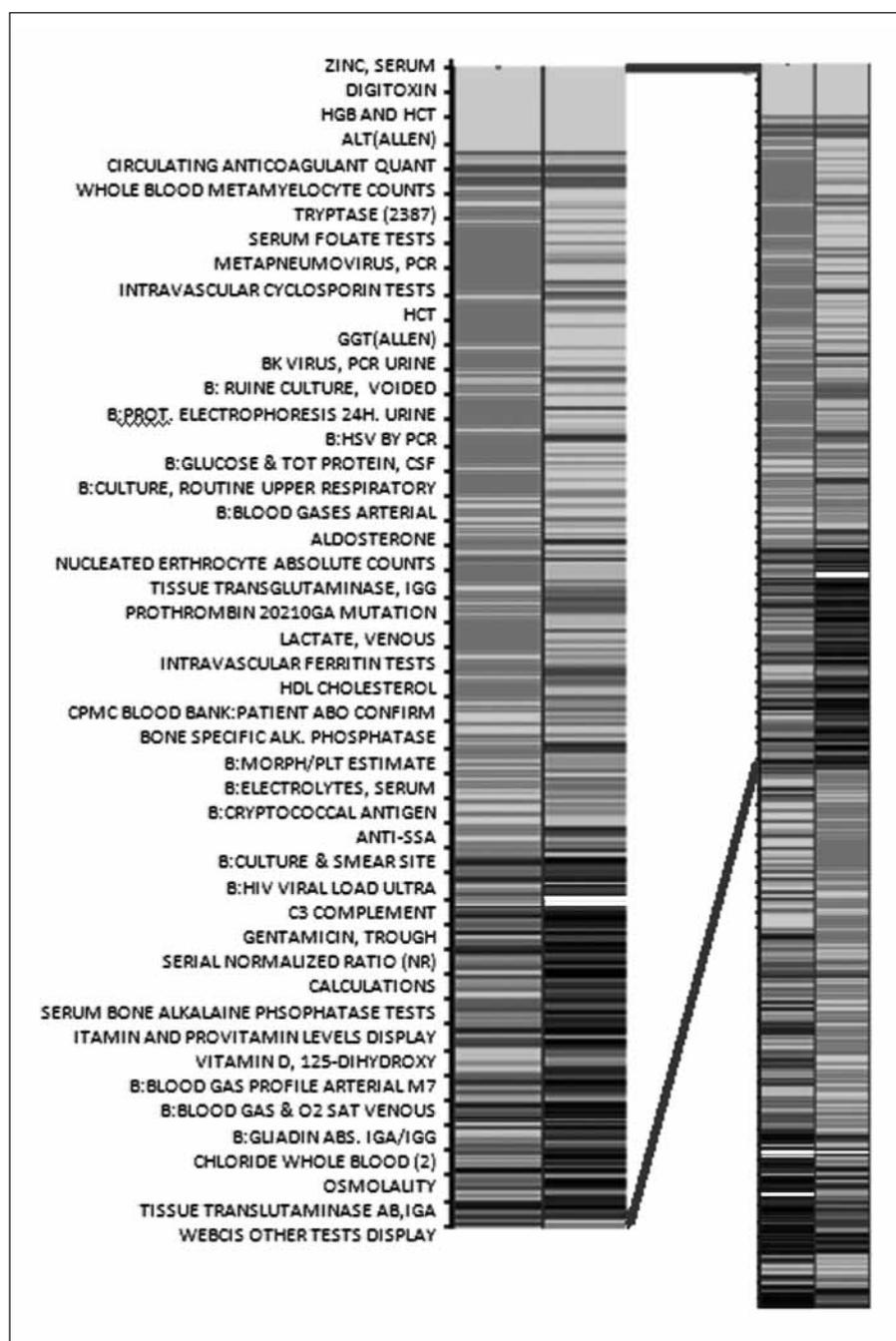
Fig. 6 Clustered test viewing, by specialty. Left column: hospitalists, right column: nephrologists. Not all test names are shown.

4.2.1 Clinician-based Analysis

Visual inspection of the resulting heatmap and its relation to the clusters shows that the method makes it simple to identify clusters of tests that are highly used by a majority of clinicians (red horizontal bands in ►Fig. 5), and those that are viewed by a small number or single clinicians (vertical lines in ►Fig. 5).

The most commonly viewed test was cholesterol 2. Other frequently viewed test results across hospitalists and nephrologists included complete blood count (white blood cell count [WBC], red blood cell count [RBC], RBC distribution width [RDW], platelet count [PLT], hemoglobin [HGB], hematocrit [HCT], mean corpuscular hemoglobin [MCH], mean corpuscular hemoglobin concentration [MCHC], mean corpuscular volume [MCV], mean platelet volume [MPV], and nucleated RBC count); MPV, serum creatinine kinase, bilirubin, phosphorus, albumin, alkaline phosphatase, uric acid, low density lipoprotein (LDL) and high density lipoprotein (HDL), hepatobiliary, blood count, and urine chemistries panels were also frequently viewed. Inspection of the individual user profiles revealed the five clinicians viewing HIV-related tests (e.g., CD3, CD4).

Clustering on the clinician axis yielded two clusters, but these did not correspond to the two specialties. One group consisted of 11 clinicians (7 nephrologists and 4 hospitalists) and the other, 12 clinicians evenly divided. This was reflected in inspection of use of individual laboratory tests: viewing of several tests such as urine chemistries were evenly distributed among hospitalists and nephrologists, as was cardiac display.

4.2.2 Specialty-based Analysis

The heatmap in ►Figure 6 shows views of data elements by nephrologists (right column) and hospitalists (left column), revealing tests common to both and specialty-specific elements.

The display simplifies comparison of test result viewing across specialties. There is considerable overlap, but also some differences.

Clustering by specialty versus laboratory tests/panels showed that, of 598 elements, 93 (15%) were viewed only by hospitalists, 130 (22%) only by nephrologists, and the rest by both specialties, with hospitalists generally viewing tests more frequently. Specialized viewing fell into expected patterns, such as nephrologists viewing parathyroid hormone, amylase, creatinine kinase, and specific urinalyses, and hospitalists viewing urine microbiology displays, blood gases, and a wider variety of tests. It also included a much higher incidence of nephrologists viewing tests for HIV and other infections, than for hospitalists. Physicians involved in renal transplant could be identified from the data by manual inspection of their viewing patterns, because of their distinctness. The re-

sults from the manual inspection were confirmed with the head of the nephrology department. One nephrologist viewed an unusual variety of seemingly unrelated tests; hypothesized to be due to his patients' participation in a research protocol.

4.2.3 Clinician/Patient/Day Analyses

In terms of comparisons between clinicians viewing the same patient records on the same day, more than 20% of matched pairs of clinicians viewed completely mutually exclusive sets of tests, indicating substantial variation in viewing patterns (▶Figs. 7 and 8).

Sixty-seven of the clinician pairs were repeated a total of 297 times during the month. Further examination of patterns among the repeated pairs might shed light on reasons for specific data selection and differences between members of the pair; however, this is beyond the scope of the current study.

Almost all (91.4%) of the clinician pairs were in the same specialty. When clinicians were in the same specialty, only 20.3% of the views were mutually exclusive, as compared to 50% when the specialties were different.

4.2.4 Analysis by ICD-9-CM Codes

▶Figure 10 illustrates the ability of our methods to distinguish data viewing patterns by ICD-9-CM code. For example, note differences between atrial fibrillation and rectal and anal ulcer test viewing. The former reflects a larger variety of less frequently viewed tests, while the latter reflects more frequent viewing of a smaller range of tests.

In summary, the clinician-based, specialty-based, clinician/patient/day, and ICD-9-CM-related analyses suggest that our methods do detect meaningful patterns. Thus, Hypothesis 2 was supported.

4.3 Hypothesis 3

▶Table 1 summarizes the test results in Eclipsys patient records and identifies the percentage (8.6%) and types of tests that

Fig. 7
Discordant elements (i.e., those not viewed by the other member of the matched pair) in laboratory test viewing among matched pairs of clinicians (n = 474 pairs) viewing the same patient record on the same day.
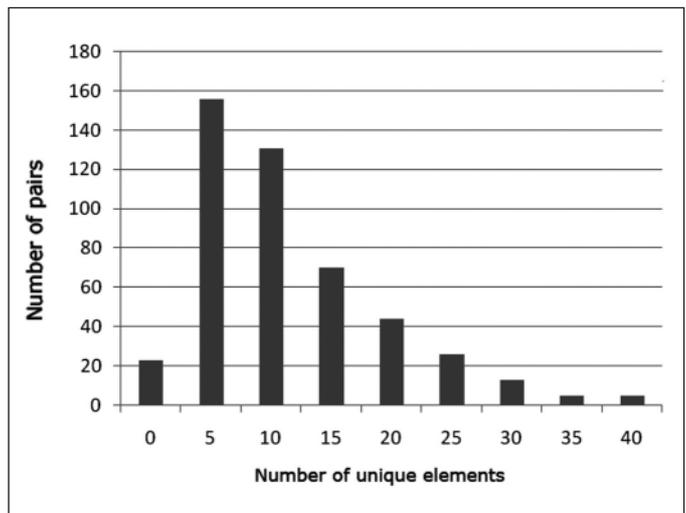
Fig. 8
Discordant elements (i.e. those not viewed by the other member of the matched pair) in laboratory test viewing among matched pairs of clinicians (n = 474 pairs) viewing the same patient record on the same day, as a percentage of total elements viewed by both clinicians.
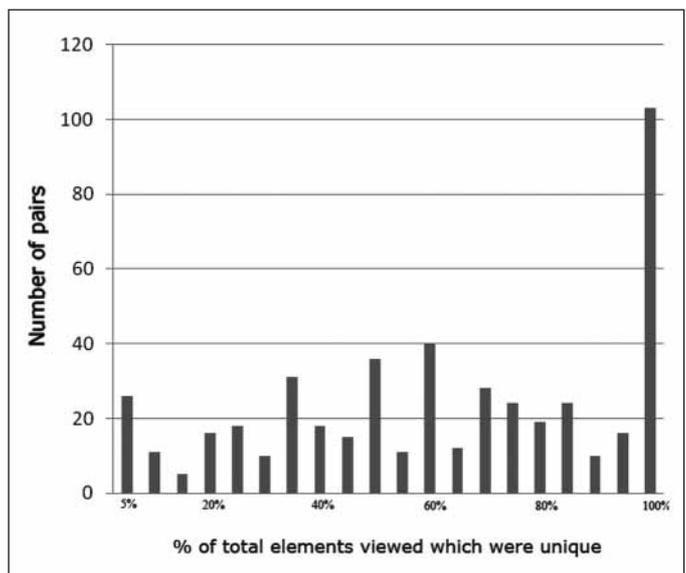
did not appear in the clusters from the heatmap and cluster analysis. For example, for Patient 2, there were 82 lab tests in the record. Our cluster-derived lab sets included 67 of these tests and missed 15; examples of the missed tests are in the 'types missing' column (8 drug screens, RPR, etc.). Thus, Hypothesis 3 (that >80% of data elements in six real patient records would be covered by clusters from our method) was supported.

5. Discussion

For our first hypothesis, the study findings supported the claim that laboratory test

viewing reflected a 'long tail' distribution where the high-frequency part of the curve corresponds roughly to the number of widgets that can easily be included in a single-view interface. While this is a somewhat arbitrary cut-off point, it corresponds to a functional distinction. This supports the rationale for user-configurable interfaces to accommodate the large variability in information needs.

In regards to Hypothesis 2, our method was able to distinguish patterns in laboratory test results viewing. The present study revealed less specialization in laboratory test viewing between the hospitalist and nephrology specialties than expected; this may be due to several factors. Firstly, this is
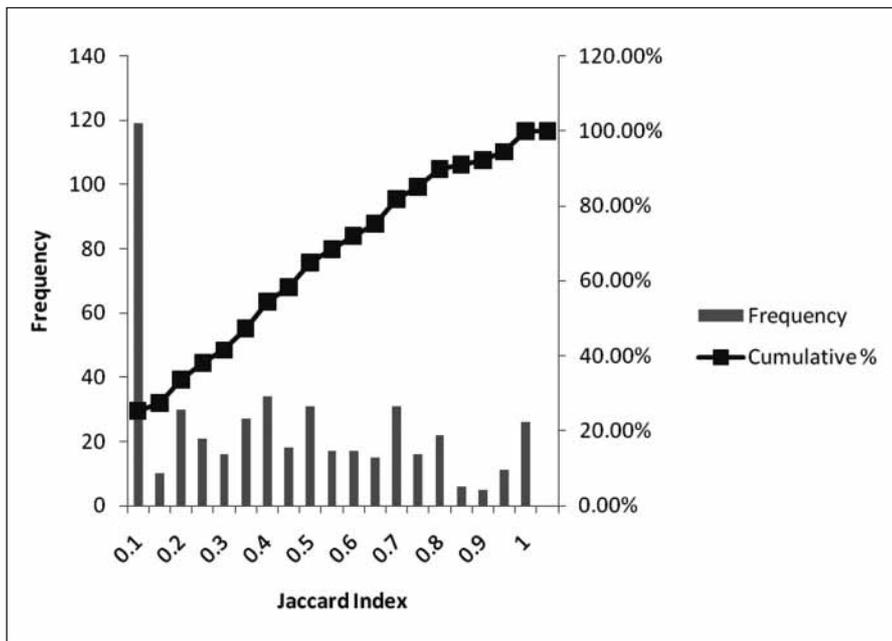
Fig. 9 Jaccard Index frequency distribution. We calculated the Jaccard index of similarity between the viewed element sets of the same matched pairs as in Figure 8.

Fig. 10 Laboratory test views by ICD-9-CM codes. The differences in testing viewing for atrial fibrillation and rectal and anal ulcer are easily visible.

a tertiary care institution in which patients at the end stages of renal failure often have multiple comorbidities, perhaps requiring that both sets of clinicians view similar information. The similar figures for numbers of different tests viewed in each specialty may be due to the fact that there may be multiple medcodes for test variations, due to institutional (facility) differences among the institutions comprising NYPH. These were not grouped together because they reflect differences in clinician behavior, need for access, and laboratory test variety, which a widget-based user interface must accommodate. The fact that clinicians clustered into two groups that did not correspond to specialty may indicate that clinician viewing patterns are driven by overall patient situation rather than just clinician specialty. This suggests that specialty-based widgets may provide a starter set of data elements, but that there is a need for further configurability. One unexpected result is the roughly equal total numbers of different tests viewed by both specialties; it might be expected that hospitalists, seeing patients with a wider range of conditions, would be viewing a wider range of tests. Since the user's specialty is known for any CIS session, the identification of specialty-specific viewing patterns would

allow us to populate the system automatically with data elements that are most frequently viewed by that specialty. This includes not only data viewing interfaces, but also population of other interface elements, such as pick lists for widgets and custom lab panel widget creation. This should save time for the user and increase task-technology fit.

Clinicians viewing records of the same patient on the same day had substantial differences in viewing frequencies for some test results. Some variation due to different access times, test result availability, and clinical progression is to be expected, but this may also reflect individual physician viewing practices. Completely different sets of tests are viewed by the pairs of clinicians more than 20% of the time; further study is needed to see how much of this is due to individual physician preference and how much is due to the other processes mentioned. While comparison of clinician pairs in the same specialty (20.3%) and different specialties (50%) reveals that there is much more similarity between those in the same specialty, as expected, the results show that even same-specialty clinician pairs viewed entirely different items 20% of the time.
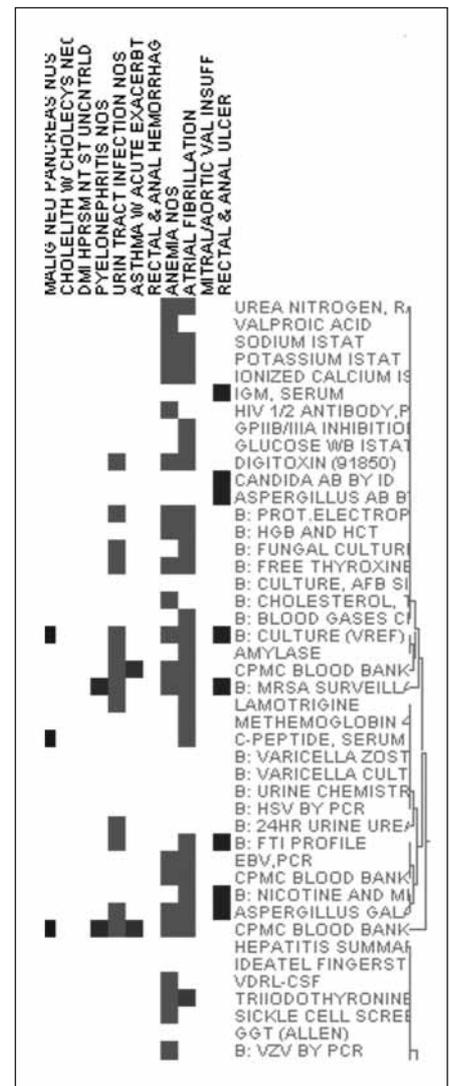
The study of ICD-9-CM codes and test result viewing must be considered with caution due to issues such as timing of code assignment and difficulty in ascertaining active and inactive problems. Nevertheless, it was possible to distinguish meaningful patterns that made sense medically, such as those in the example (rectal and anal ulcer and atrial fibrillation). If confirmed in future studies, this supports the possibility of diagnosis-relevant tests as widget collections.

In terms of our third hypothesis, the inspection of Eclipsys records of 30 typical nephrology and general inpatients revealed

that the clustering-derived sets captured most of these patients' laboratory tests; with only 51 not present in 589 test results. Although the sample is small and the finding must be confirmed in a larger sample, this suggests that widget collections pre-populated by our methods are likely to include the great majority of tests available for clinicians to view.

The ideal for using these studies to formulate widget collections is to maximize efficiency by providing every test the clinician may need to view while minimizing wasted space and clinician attention by not including unnecessary tests. The consequences of inclusion of extras are not severe, however, as low-priority tests can be placed at the bottom of the page; indeed, one use of our study results can be to prioritize placement (e.g., upper left corner for highly viewed elements).

Limitations of this study include the 1-month length of time for which logfiles were analyzed; this may reflect seasonal variation and presents a smaller sample size. Analysis by panel also differentiated tests which may be identical but bear different names because of use at different institutions; this is a very small proportion of tests. False positives, i.e. panels that are identified as being relevant but are actually not used in context, were not enumerated in this study, but are relevant for future exploration. The consequences of false positives are less severe than for false negatives (i.e., omission of a needed element). Inclusion of elements which are not needed may decrease efficiency but does not have the potential to cause errors of omission the way false negatives would. Only hospitalists and nephrologists were studied; thus, patterns of distinction may not be reflective of other medical specialties.

6. Conclusion

The method proved useful in rapidly and succinctly identifying patterns of information viewing among clinicians, and particularly which information is very frequently viewed by most clinicians and which is viewed by a few individuals or subgroups in specialized patterns. This supports the feasibility of our methods to

inform development of initial sets of widgets for the widget-based interface. The fact that distinct patterns were found for individual users (due to role, involvement in different activities (e.g., transplant), case load, or personal preferences confirms the need for the functionalities afforded by user-configurable information access. The

Table 1 Comparison between laboratory tests in 30 Eclipsys patient records and clusters derived from heatmap and cluster analysis

Patient	#Tests in Patient Record	#Tests in Patient Record and Not in Clusters	Types of Tests in Patient Record and Not in Clusters
1	67	0	
2	82	15	8 drug screens, RPR (rapid plasma reagin), protein electrophoresis, C3 and C4 complement
3	147	3	Tobramycin trough, Vancomycin trough, Clostridium difficile
4	52	3	Hepatitis B surface and core Antigen, Hepatitis B Qualitative
5	31	0	
6	38	1	Digoxin
7	3	0	
8	6	0	
9	7	1	C. difficile toxin
10	7	1	Phosphorus
11	11	2	Ddimer, fibrinogen
12	4	0	
13	4	1	Phosphorus
14	11	1	Drug screen, serum amylase
15	9	1	Phosphorus
16	9	2	Phosphorus, Free Phenytoin level
17	7	1	Phosphorus
18	9	2	Phosphurus, serum amylase
19	6	1	Phosphorus
20	7	1	Phosphorus
21	9	2	vancomycin trough, Phosphorus
22	7	1	Phosphorus
23	7	1	Phosphorus
24	11	3	CSF cell count, CSF smear, CSF glucose + total protein
25	7	2	Phosphorus, Free Phenytoin level
26	5	1	Phosphorus
27	5	1	Enteric pathogens – stool
28	4	0	
29	11	3	Creatine Phosphokinase,P,
30	6	1	B Type Natriuretic Peptide
Total	**589**	**51**	

methods that we have described may be useful for others seeking automated approaches to inform widget-based interfaces for CIS. In addition, they contribute to our general understanding of how clinicians use CISs.

References

1. Senathirajah Y, Bakken S. A User-Configurable EHR using Web 2.0 Approaches. AMIA Spring Congress. Phoenix, AZ: AMIA; 2008.
2. Senathirajah Y, Bakken S. Architectural and Usability Considerations in the Development of a Web 2.0-based EHR. Advances in Information Technology and Communication in Health. IOS Press; 2009. pp 315–321.
3. Senathirajah Y, Bakken, S. Rapid configuration of a widget-based EHR interface to meet an emerging need. Public Health Informatics Pre-meeting, NLM Trainee Conference, Portland, OR; 2009.
4. Senathirajah Y, Bakken S. Development of User-configurable Information Source Pages for Medical Information Retrieval. AMIA 2007 Symposium Proceedings Page 2007: 1109.
5. Chen ES, Cimino JJ. Patterns of usage for a Web-based clinical information system. Stud Health Technol Inform 2004; 107 (Pt 1): 18–22.
6. Chen E. Knowledge discovery in clinical information system log files and the implications for wireless handheld clinical applications (Doctoral dissertation). New York: Columbia University; 2004.
7. Anderson C. The Long Tail: Why the Future of Business is Selling Less of More. Wired; 2004.
8. Anderson C. The Long Tail: Why the Future of Business is Selling Less of More. New York: Hyperion; 2008.
9. Brynjolfsson E, Hu Y, Smith M. From Niches to Riches: Anatomy of the Long Tail. MIT Sloan Management Review 2006; 47 (4): 67–71.
10. Hripcsak G, Cimino JJ, Sengupta S. WebCIS: Large Scale Deployment of a Web-based Clinical Information System. AMIA Annu Symp Proc 1999. pp 804–808.
11. Cimino J. From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. J Am Med Inform Assoc 2007; 7: 288–297.
12. Wright A, Sittig D. Automated Development of Order Sets and Corollary Orders by Data Mining in an Ambulatory Computerized Physician Order Entry System. AMIA Annu Symp Proc 2006. pp 819–823.
13. Kriegel H, Kröger P, Zimek A. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM Transactions on Knowledge Discovery from Data 2009; 3 (1): 1–58.
14. Eisen M, Spellman P, Brown P, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998; 8: 14863–14868.
15. Golub T, Slonim, DK, Tamayo, P, Huard, C, Gaasenbeek, M, Mesirov, JP et al. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science. 1999; 286: 531–537.
16. Madeira S, Oliveira A. Biclustering Algorithms for Biological Data Analysis: A Survey. IEEE/ACM Transactions on Computational Biology and Bioinformatics 2004; 1: 24–45.
17. Quackenbush J. Computational analysis of microarray data. Nat Rev Genet 2001; 2: 418–427.
18. Madeira S, Oliveira A. Biclustering Algorithms for Biological Data Analysis: A Survey. IEEE/ACM Transactions on Computational Biology and Bioinformatics 2004; 1 (1): 24–45.
19. Auman JT, Boorman GA, Wilson RE, Travlos GS, Paules RS. Heat map visualization of high-density clinical chemistry data. Physiol Genomics 2007; 31 (2): 352–356.
20. Melcher J, Seese D. Visualization and Clustering of Business Process Collections Based on Process Metric Values. 10th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing 2008. pp 572–575.
21. Moona J, Hyun-Jin J, Moon MH, Chunga BC, Choia MH. Heat-Map Visualization of Gas Chromatography-Mass Spectrometry Based Quantitative Signatures on Steroid Metabolism. Journal of the American Society for Mass Spectrometry 2009; 20 (9): 1626–1637.
22. Datta S, Datta S. Evaluation of clustering algorithms for gene expression data. BMC Bioinformatics 2006; 7(Suppl 4): 17.
23. Datta S, Datta S. Comparisons and validation of statistical clustering techniques for microarray gene expression data. Bioinformatics 2003; 19 (4): 459–466.
24. Mobasher B, Dai H, Sun Y, Zhu J. Integrating Web Usage and Content Mining for More Effective Personalization. Web Mining for E-commerce Workshop (WebKDD'2000) in conjunction with ACM-SIGKDD Conference on Knowledge Discovery in Databases2000.
25. Gemmell J, Sehepitsen A, Mobasher B, Burke R. Personalization in Folksonomies Based on Tag Clustering. 6th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems, held at AAAI; Chicago, USA; 2008.
26. Horsky J, Kaufman, Patel, VL. The Cognitive Complexity of a Provider Order Entry Interface. AMIA Annu Symp Proc 2003. pp 294–298.
27. Horsky J, Kaufman D, Oppenheim M, Patel V. A framework for analyzing the cognitive complexity of computer-assisted clinical ordering. Journal of Biomedical Informatics. 2003; 36 (1–2): 4–22.
28. Patel V, Kushniruk A. Cognitive and usability engineering methods for the evaluation of clinical information systems. Journal of Biomedical Informatics 2004; 37 (1): 56–76.
29. Seo J. Hierarchical Clustering Explorer. 2004. Available from: http://www.cs.umd.edu/hcil/hce/
30. Ho Y. Heuristics, Rules of Thumb, and the 80/20 Proposition. IEEE Transactions on Automatic Control 1994; 39 (5): 1025–1026.