AA Software Development.

Data Selection Error Prevention

Date: 22 Feb. 2008

Author: Arie Avnur

I. General Overview

This documents and describes a mechanism for reducing data selection user error rates. Application: Clinical data records management systems like the cobas IT solutions and potentially the software that controls analyzers, their data management systems and work area management systems.

A couple of cases (PCC's) relating to user error in selecting a record, mostly wrong patient assigned to an order, created the motivation to challenge the current state on the subject.

The point here is that users expect the system to help them make the right selection, or prevent them from making the wrong one!!!

II. Problem Definition

II-A. Errors in data selection.

In many cases, users of data management systems select records from system repository. Typically, this is done by selecting a record from a list. The "candidates" list is either all the records of the right type to be selected or a result of a search or filtering.

An example can be the selection of a patient to be assigned to an order. It is a common task, and a critical one. Selecting and assigning the wrong patient to an order creates a potentially dangerous situation.

In many cases similar records appear next to each other in the selection list. Especially since most lists are sorted by name or similar attributes. It is then not inconceivable that the user might point and select the wrong record. This could happen as a result of **environmental conditions** (like task and time pressure), **user ability** to discriminate between similar lines (user specific performance), and the **similarity among records**.

Error rates

The following diagram is an example of a couple of records so similar, that they pose or increase the risk of wrong selection:

Patient ID.	Last Name	First Name	Age	Gen
1234567	David	Larry	55	M
1936022	Doe	John	21	M
1936827	Doe	Jeff	54	M
3958133	Eckerman	John	34	M

III. Discussion

III-A. Records Representation Distinction

The root cause of this class of errors is similarity between records and their representation for selection. E.g. patients with the same last name and the same address are quite common for obvious reasons. In some cases their gender will be identical, and their first name similar enough, increasing the probability of error in distinguishing among them. To measure or assess the problem or similarity, it is proposed to use the concept of ***record distinction*** as measured by the ***distance*** between two

records. We can define the distance between two records as: The number of different characters out of the total characters making up record's data[1]. It includes the text representing all of records' attributes. This however might not be the best or true measure to characterize the problem. Attributes that are different between records might not be displayed on the selection list. E.g. if the list represented patients only by their last name and first name initial, we could end up with many records that look identical by their representations and it will be very difficult to select among them.

We therefore introduce a second measure: ***representation distance***—the distance (or number of different characters) between records' **representation**—the string or strings representing the record for selection. Naturally, the more the representation distance between two records, the easier it is to distinguish between them and select the correct record among them.

III-B. Representation Distance and Error Rates

The following equation is an attempt to model selection error rate qualitatively based on its main factors:

SER = 1 / (K * RD/RT * DH * UC)

where:

SER stands for Selection Error Rate and should be minimized,

K is a proportion factor,

RD is the representation distance between the records, in characters,

RT is the total representation string length

DH is the level that distance is highlighted. DH = 1 if the differentiating parts of the records' representation are displayed with the same attributes as the equal parts or as the rest of the selection list. Dh > 1 if the discriminating text is highlighted in a way that helps the user distinguish between the records and identify the one the user is looking for, and

UC is a user dependent factor; some users will tend to make more errors than others under the same conditions.

From this equation it is apparent that in order to reduce error rate we should:

a. increase representation distance of list items,
b. maximize the representation distance ratio to total representation length (e.g. to record names, differing from each other by two characters, having total length of 4 characters, are way less likely to be mistakenly selected than if their name representation was 20 characters long).
c. Highlight the difference portion of the record representation
d. Train users in selecting from system lists or menus, or screen for those users who make less mistakes.

IV. Proposed Solutions

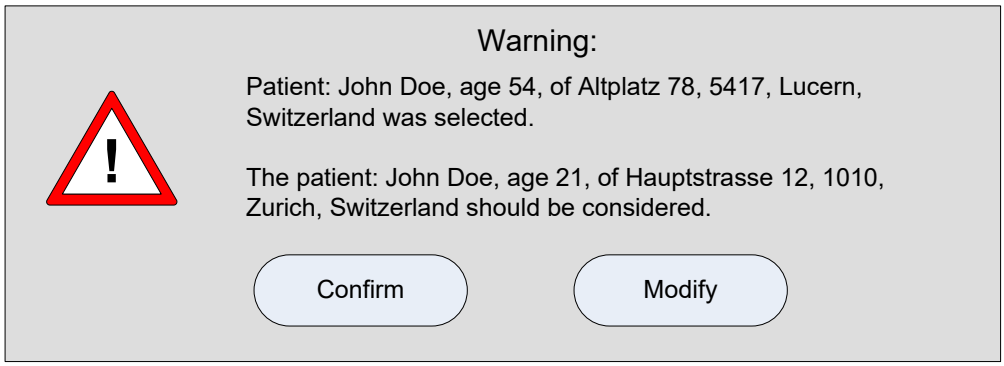
Following is a list of several measures that can reduce record selection error rates by increasing record representation distance and its highlighting. These possible measures are generally listed from the easy to the more complex to implement.

IV-A. Record Selection Warnings

Upon record selection, the system checks if the selected record is among records with similar representation, and posts a confirmation prompt for the selection. This prompt may present and highlight the differences between similar candidates:

---

[1] This definition is borrowed from the field of error detection and correction codes in data communications.



IV-B. Highlighted Record Differentiation Representation

The system highlights the differentiating parts of record representation in the selection list by displaying it with special attributes like bold, font size, italics, contrast color, background color, etc.

Patient ID.	Last Name	First Name	Age	Gen
1234567	David	Larry	55	M
1936022	Doe	**John**	**21**	M
1936827	Doe	**Jeff**	**54**	M
3958133	Eckerman	John	34	M

IV-C. Dynamic Record Representation

The system adds differentiating data to similar records:

Patient ID.	Last Name	First Name	Age	Gen	Additional data
1234567	David	Larry	55	M	
1936022	Doe	John	21	M	Hauptstrasse 12, 1010, Zurich, Switzerland
1936827	Doe	Jeff	54	M	Altplatz 78, 4517, Lucern, Switzerland
3958133	Eckerman	John	34	M	

Or better yet:

Patient ID.	Last Name	First Name	Age	Gen	Additional data
1234567	David	Larry	55	M	
1936022	Doe	John	**21**	M	Hauptstrasse 12, 1010, Zurich, Switzerland
1936827	Doe	John	**54**	M	Altplatz 78, 4517, Lucern, Switzerland
3958133	Eckerman	John	34	M	

IV-D. Downstream Verification

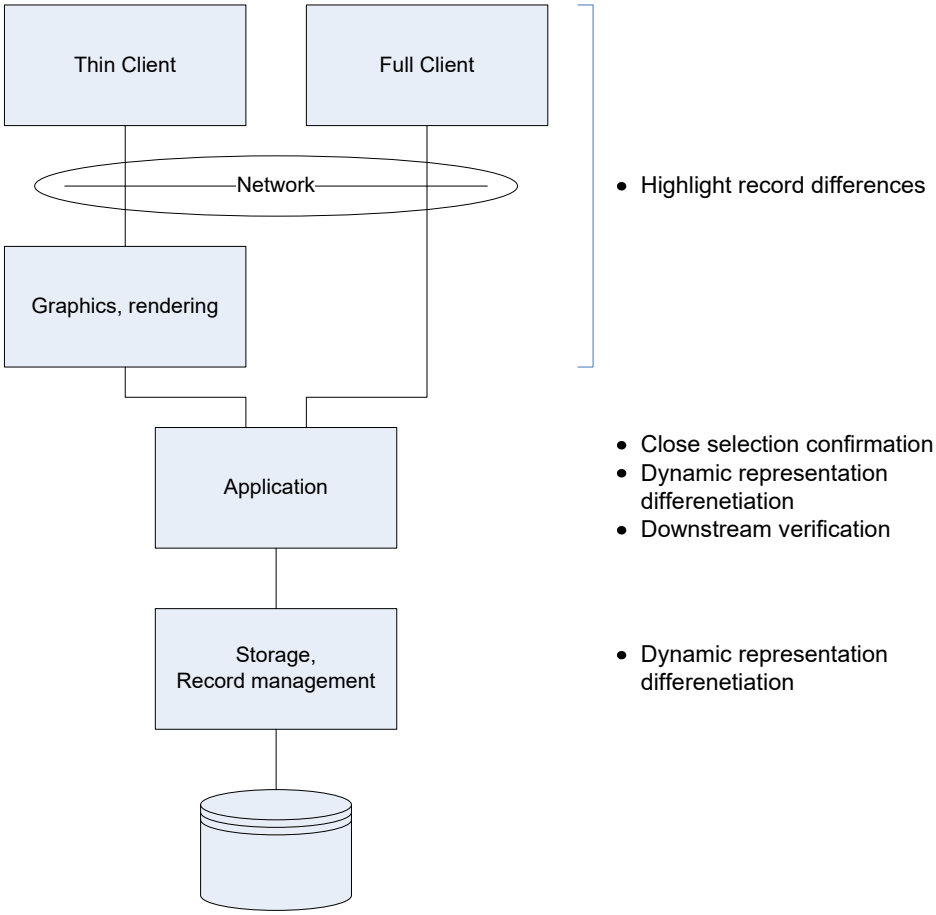
What if down the road, a test result is reported that doesn't quite agree with some of patient's demographics, like his age or gender? It could point to error in patient selection when the order was created. A couple of actions can be taken to remedy this situation:

a. Record the confidence level of the selection. If the record was selected out of very different ones, like in a normal case, it will be assigned a high confidence level. If it was selected from a few similar records, its confidence level is lower. The equation in III-B above could be used for that purpose. When a data item conflicting with low confidence level selection is encountered, a warning may be issued for the user to reconsider the selection.

b. Record the unselected “close” records choices. Later on, when conflicting data is encountered, it can be checked against those options, and the more likely selection may be offered for reconsideration.

IV-E. Implementation Considerations

The features described here may make the system look smarter, but are not in the realm of artificial intelligence. Relatively simple algorithms are required that measure string differences or distances, and add more differentiating data to record representation if its neighbors are too close. Downstream selection verification is somewhat more complicated though. It requires special records, storing selection confidence levels and unselected “candidates”, as well as logic that uses this data to inform of or suggest alternatives to the user. The following diagram depicts a multi-tier generic system’s layers and suggests which of the feature can be implemented at which layer.



V. Summary

Errors in selecting from a menu are a commonplace and more so when menus contain similar entries. The proposed mechanism can warn users when selection is made among such similar records, allowing the user to focus and reduce selection error rates. Warning messages and popups that appear repeatedly and predictably become less effective as users get use to them. A dynamic mechanism is proposed that pops up warnings when actual data create higher error risk. Such error warning should also highlight the possible error, making them less routine and user friendlier.

Any combination of the features described above has the potential of reducing record selection error rates.