Data migration of consistent Sickle Cell Disease clinical-database at Muhimbili National Hospital in Tanzania

Raphael Z. Sangeda[1], Hans Morsch[1], Evarist Msaki[1], Frederick Mbuya[1], Sharon Cox[1], Amel Ghouila[2], Bruno Mmbando[1], Alia Benkahla[2], Charles Newton[1,3], Kais Ghedira[2], Julie Makani[1]

[1] Muhimbili University of Health and Allied Science, Dar es Salaam, Tanzania.

[2] L'Institut Pasteur de Tunis (IPT), Tunis, Tunisia.

[3] KEMRI-Wellcome Trust Research Program CGMRC, Kilifi, Kenya.

Abstract

Background

Proper usage of biomedical data can be useful in designing future research strategies and in patient management and care. Exploring the clinical data with text mining allow the prediction of healthcare costs, disease diagnosis and prognosis, and the discovery related healthcare pat terns from other databases establishing, relationships among diseases, and relationships among drugs. Moreover; clinical research on the data can allow discovering disease intervention strategies.

However, exploration of clinical data can be hindered if the database is not properly designed, un- normalized or contain redundant data. We hereby describe the automated migration of a clinical database containing cohort data of Sickle cell disease (SCD) patients attending clinic at Muhimbili National Hospital, in Dar es Salaam, Tanzania. The main objective was to migrate in the old SCD database into a new consistent database.

Methodology

The new database was modeled using MySQL-Workbench by consulting the study case report forms (CRF) and the old database.

Mapping of new database columns was matched with the new database columns to get a definitive column for migration. Values in the old database were checked against valid range of values and text values were recorded to uniform values for all clinical variables. Consistency check within and between tables was done and data cleaned. The definitive column mappings were made and these were used to guide the automated migration from old to new database.

Consistency, range check and mapping were done using python scripts or VBA scripts. The input to the script was made via comma separated values (csv) files.

Results

In total 1768 columns in 33 tables of the old database were considered important to migrate into 622 columns within 27 new tables in the new database. Nine pairs out of the 33 tables were double entry tables, which were merged in the new database design. The resulting mapping comprised of 1712 column pairs where related old columns were migrated to the same new column.

Conclusions

The newly created repository will facilitate analyzing the clinical data and generate new interventions for SCD

Keywords — Data migration, Sickle Cell Disease, clinical-database