**Review - Potter’s Wheel: An Interactive Data Cleaning System**

Vijayshankar Raman and Joseph M. Hellerstein

**Summary:** The paper presents an interactive, user friendly data cleaning system that integrates transformation and discrepancy detection: Potter’s Wheel. Potter’s wheel system save user from writing hard complex codes by making use of spreadsheet like interface enabling users to execute real time transformations using simple graphical operations while displaying the effects on real data.

**Introduction:** In today’s electronic era, where computers are used extensively, one of the main challenge is cleaning thelarge amounts of untidy data and transforming it into clean uniform format. Existing approaches are inefficient in the terms of transformation time and the effort that is put in designing and execution of long complex codes over huge amounts of data.

**Literature Review:** The task of data cleaning goes through three steps: auditing, transformation selection, and transformation application. Existing market solutions come in two forms: Auditing tools like Unitech Systems' ACR that detect discrepancies, and Data transformation tools like Data Junction or DataStage that transform data in the desired format. Although, these tools are designed to handle large datasets but still are far from being efficient to accurately and timely transform the data. Furthermore, these systems require complex coding tasks, making it difficult for users.

**Methods:**Potter's wheel simplifies the tedious task of data cleaning by integrating transformation and discrepancy detection under one roof. The described system presents an interactive spreadsheet-like interface for the user to build transformations graphically. The effects of the transformations are shown in real time, making it easier to rollback. Additionally, the system executes discrepancy detection automatically in the background. Potter's wheel makes data cleaning a tight, closed-loop process where users can gradually develop and refine transformations as discrepancies are found. Strings are parsed using user defined domain structures that results in a general and extensible discrepancy detection mechanism.

**Results:** Authors described Potter’s Wheel, an interactive system for data transformation and cleaning. By integrating discrepancy detection and transformation, Potter’s Wheel allows users to gradually build a transformation to clean the data by adding transforms as discrepancies are detected. Users can specify transforms through graphical operations or through examples, and see the effect instantaneously, thereby allowing easy experimentation with different transforms.

**Discussion:** Although, the presented system makes it easier for user to do the transformation, however, the efficiency of this interactive user interface is not measured. Moreover, the system only deals with the text data and no other types of data. Last but not the least, the automatic discrepancy detection depends upon the scrolling by user.

**Conclusion:** Potter’s wheel is a user-friendly data cleaning system that integrates transformation and discrepancy detection. Potter’s wheel system saves user time by eliminating hard complex coding by using easy to use interface enabling users to execute real time transformations using simple graphical operations while displaying the effects on real data.