TECHNOLOGY

Data Cleaning Systems

OVERVIEW

There is increasing interest in methods for computer-aided rapid data cleaning. These methods deal with detecting and removing errors and inconsistencies from data in order to improve data quality.

Researchers at the University of Maryland have developed six methods for flagging statistical anomalies within the textual content of electronic dictionaries stored in the XML format. The methods can be classified into single-field methods and tied-field methods. Single-field methods detect anomalies on the basis of expectations inferred from content in the same single field type. Tied-field methods detect anomalies on the basis of expectations of content correspondences inferred from content in multiple related fields.

CONTACT INFO

UM Ventures
0134 Lee Building
7809 Regents Drive
College Park, MD 20742
Email: umdtechtransfer@umd.edu
Phone: (301) 405-3947 | Fax: (301) 314-9502

Additional Information

INSTITUTION

University of Maryland, College Park

PATENT STATUS

Pending

EXTERNAL RESOURCES

IS-2015-175