Appendix of Data Cleaning tools

Toped++

Bellman [5]

SWYN [2]

Potluck [14] applies simultaneous text editing [19] to merge data sources.

Karma [26] infers text extractors and transformations for web data from examples entered in a table.

Vegemite [18] applies PBD to integrate web data, automates the use of web services, and generates shareable scripts.

Other interfaces [15] apply PBD to data integration via copy and paste actions.

Potter’s Wheel [22] provides a transformation language for data formatting and outlier detection.

Ajax [8] also provides an interface to specify transforms, with advanced facilities for entity resolution.

Google Reﬁne [13] (formerly Freebase GridWorks) leverages Freebase to enable entity resolution and discrepancy detection.

The above are extracted from <http://vis.stanford.edu/files/2011-Wrangler-CHI.pdf>

(I think Yau’s first book has a pretty extensive list as well).