Cognitive Dimensions of Between-Table Context Support in Direct Manipulation Wrangling Interfaces.

Steve Kasica
Nov. 19, 2019
Workflows to reproduce
Medicaid long-term managed care reports

• Comes from a CAR workshop / Columbia course tutorial from Sara Cohen. [Cohen 2014]

• Data is New York’s Medicaid long-term managed care reports.

• In the taxonomy
  • This kind of wrangling falls under Tidy Up Presentation Data
  • This workflow performs the following wrangling tasks:
    • Extract date from string
    • Extract entity from string
    • Aggregate join (a good task for between-table context)
Cognitive Dimensions
Notational Dimensions

• There are 13 different dimensions.
• A common vocabulary to discuss usability in user interfaces
Viscosity

• “Resistance to change” [Blackwell et al, 2003]
Visibility

• “Ability to view components easily” [Blackwell et al., 2003]
  • Can we see all components in VPL? [Blackwell et al., 2003; Green, 1996]

• In data wrangling, visibility because an issue as datasets become large
  • Is every part of the relevant data simultaneous visible?
  • In high-dimensional data you have to scroll to view all columns
  • In data with many observations, you have to scroll to view rows.
  • **Focal point:** Would increasing visibility may decrease error-proneness?
  • Visualization may help here. Charts are great at representing data compactly, a.k.a data-ink ratio [Tufte, 1983]
Premature Commitment

• “Constraints on the order of doing things” [Blackwell et al., 2003]
Hidden dependencies

• “Important links between entities are not visible” [Blackwell et al., 2003]
Role-Expressiveness

• “The purpose of an entity is readily inferred” [Blackwell et al., 2003]
• In data wrangling, it is already difficult to verbally express table transformations.
• Different tools use different vocabulary to describe the same thing.
  • Entity resolution: “cluster and edit” and “mass edit” in OpenRefine and “standardize” in DataPrep
  • DataPrep does include little icons, which are more helpful than no icons.
Error-Proneness

• “The notation invites mistakes and the system gives little protection.” [Blackwell et al, 2003]

• In data wrangling, errors often creep in when filtering as Type I vs Type II errors in the gulf of execution and evaluation [Hutchins et al., 1985]
  • Type I / false positive: A row was removed, but it should have been kept.
  • Type II / false negative: A row was kept, but it should have been removed.

• You often have to approve operations on rows that you don’t know the values of.
Abstraction

• “Types and availability of abstraction mechanisms” [Blackwell et al, 2003]
Secondary notation

- “Extra information in means other than formal syntax” [Blackwell et al, 2003]
Closeness of mapping

• “Closeness of representation to domain” [Blackwell et al, 2003]
Consistency

• “Similar semantics are expressed in similar syntactic forms” [Blackwell et al, 2003]
Diffuseness

- “Verbosity of language” [Blackwell et al, 2003]
Provisionality

• “Degree of commitment to actions or marks” [Blackwell et al, 2003]
Progressive evaluation

• “Work-to-date can be checked at any time” [Blackwell et al, 2003]